UNDERSTANDING POPULATION, GENETIC AND ANTIGENIC DIVERSITY OF THE POULTRY RED MITE (*DERMANYSSUS GALLINAE*) TO AID DEVELOPMENT OF NOVEL METHODS FOR CONTROL

Eleanor Ivy Karp-Tatham

Supervisors: Prof. Damer Blake, Prof. Fiona Tomley, Dr Dong Xia and Dr Alasdair Nisbet

Thesis submitted for the degree of Doctor of Philosophy

December 2020

"The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information."

STUDENT DECLARATION FORM

On submitting the thesis, a student will be required to insert the following signed Declaration into their bound thesis:

I certify that:

1. The thesis being submitted for examination is my own account of my own research;
2. My research has been conducted ethically;
3. The data and results presented are the genuine data and results actually obtained by me during the conduct of the research;
4. Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis;
5. Where any collaboration has taken place with other researchers, I have clearly stated in the thesis my own personal share in the investigation;
6. The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination;
7. The thesis submitted is within the required word limit as specified by the RVC; the total number of words is given below.

Student signature: Eleanor Karp-Tatham Date: 02-12-2020

Word count: 74,713

ABSTRACT

The poultry red mite (*Dermanyssus gallinae*), an obligatory blood feeding ectoparasite, is primarily associated with poultry where it is predicted to incur losses of ~€230 million per annum from European farmers. Current control strategies, including the use of acaricides and desiccant dusts, are mostly ineffective and widespread resistance to acaricides has been demonstrated across Europe. Alternative methods of control are urgently required for *D. gallinae* and methods under investigation include the development of a suitable vaccine. One major consideration for the development of a vaccine is the extent and rate of occurrence of genetic diversity within mite populations. This project aimed to gain an understanding of *D. gallinae* population structure and antigenic diversity relating to development of novel control strategies. Genetic diversity was studied on both an inter- and intra-farm level, within the UK and across Europe. *Dermanyssus gallinae* isolates were sampled from 25 UK farms and 84 mainland European farms, spanning 16 countries. Phylogenetic analysis of the cytochrome C oxidase subunit I gene revealed admixture between isolates from the UK and the rest of the world, highlighting the occurrence of intra-farm diversity and differences in genetic diversity between production systems. The genome analysis toolkit best practices pipeline for single nucleotide polymorphism (SNP) and insertion/deletion variant calling was modified to be self-validating and utilised for identification of 32,599 SNPs through comparison of transcriptomic read sets and the *D. gallinae* genome assembly. Analysis of 145 SNP markers from 75 pooled *D. gallinae* samples collected from across the UK and Europe showed high spatial genetic diversity with significant linkage disequilibrium. Widespread occurrence of mutations relating to pyrethroid resistance located in the voltage-gate sodium channel was demonstrated in *D. gallinae* populations through nine putatively associated markers. Investigation into antigenic variation in four vaccine candidates demonstrated negative dN/dS ratios and variable levels of diversity.

ACKNOWLEDGEMENTS

Firstly, and foremost, I wish to thank Prof. Damer Blake, for four years of exceptional supervision, support, patience, guidance and dealing with my endless 'data-dumping' and 'dump it Fridays'. To Prof. Fiona Tomley, for always having a kind word to say whenever it was needed and belief in my abilities, even when I lack my own. To Dr Dong Xia, for having the patience of a saint when helping me to achieve the GATK pipeline and every mental breakdown involved in the process. To Dr Alasdair Nisbet for surviving countless 'just seeing if there was an update on the genome, too many 'too long didn't read' update emails. I wish to acknowledge all four supervisors for their caring attitude and support during numerous periods of ill-health.

To my wonderful parents, Angela Karp and Arthur Tatham for a lifetime of encouragement, support, endless advice and always being there. To my lovely brother, Marcus Karp-Tatham, for always being proud, believing in me, countless of hours building useful furniture for my flat and always providing a good laugh when it was needed.

To all the lovely friends in my life who have supported me through the highs and lows of the last four years, including Laura Evans, Kirsty Laird, Jack Caithness, Michael Easter, Tom Rhoden-Farmer, Thomas Bliss, Jasper-Bell Osborne, the Lincolnites and many others. With special thanks to Tatiana Küster for being an incredible colleague, a loving friend and being a second mother to me, for providing endless knowledge and advice, hours of mite sampling and chicken work made indefinitely better through her companionship, helping me through my first conference talk and unwavering faith. Finally, for being a continual source of inspiration in my life and pushing me to keep reaching for my dreams. To Kate Dulwich, for years of loyal and loving friendship, seeing me through endless emotional outbursts, healthy competition to push myself, always having the right thing to say, all the coffee and walks, always finding the time to be there and being a fellow PhD student, I am as proud of you as you are of me. From undergraduate, to masters, to PhD. We did it! To Fiona Dale, for being the best part of PhD induction and being the kind of flat mate and 'girlfriend' that words cannot do justice. Thank you for four years of enduring every good and bad day, for dealing with every failed laboratory experiment, for every angry Ellie moment, every sad Ellie moment, every happy Ellie moment and every moment in between and for all the hours you waited in hospital waiting rooms and for always taking care of me when I have been sick. Lastly, for surviving lockdown I together, all the exercise we've enjoyed together and all our dates, we made it to the finish line! To Rachel Harron, for being a great friend, for inspiring me to take up running and endless hours listening to me moan whilst we did so, being a coffee companion, picking me up on the days when I couldn't carry myself,

always making me laugh and for being the sole reason sanity was maintained during lockdown II. To David Arnfield, for on-campus support of an excessive number of rants and moans, always with positivity and cute birthday songs. To Simon Rogerson, for enduring years of my friendship, endless rants constantly, always having time and a kind word to say, never failing to put faith in my abilities and providing the best hugs. To Reece Cain, for providing light in my life during darkness, endless kindness and support, providing support during the toughest stage of my PhD and helping alleviate stress during high pressure periods.

A thank you to the RVC Parasitology group, past and present: Michelle Jones, Matt Nolan, Hanna Wickenden, Francesca Soutter, Sungwon Kim, Kim Harmen, Daniel Ngugi, Sarah McDonald, Ivan-Pastor Fernandez, Virginia Hernandez, Kelsilandia Aguiar-Martins and Gonzalo Arsuaga Sanchez. With extended thanks to Fran, Shelly and Lizzie for providing all the gossip, support in the lab during difficult times, providing comfort during many episodes of crying and enjoying a vast amount of coffee and cake together.

To the team at the Moredun Institute, Katherine Bartley and Francesca Nunn for welcoming me at the Moredun Institute and providing me with additional knowledge relating to PRM and to Dan Price for venturing into a collaboration.

An acknowledgement to all the UK farmers who kindly sent samples and questionnaire data to aid in sampling, with a special note of appreciation to Grassington Farm, Kings Farm and Buildings Farm for allowing me on-site access to their farms throughout the course of my PhD. Extended thanks to the following for contribution of European samples and all colleagues through COREMI, including: Thomas Van Leeuwen (Belgium), Teresa Mateus (Portugal), Veli YilgÇr Sirak (Turkey), Monique Mul (the Netherlands), Cristian Magdas (Romania), Susanne Kabell (Denmark), Lise Roy (France), Miroslav Radeski (Macedonia), Tana Shtylla Kika (Albania), Christina Strube (Germany), Antonio Camarda (Italy), Danijela Horvatek Tomic (Croatia), Martina Lichovnikova (Czech Republic) and José Francisco Lima Barbero (Spain). With special thanks to Prof. Thomas Van Leeuwen for forming a collaboration relating to acaricides and making it an enjoyable experience.

To the Eurofins team, who helped guide me through the development of the SNP genotyping assay, dealt with the endless complications involved and had patience to see it through.

Lastly, I wish to thank the BEMB for providing the funding, without which this project would not have been possible, as well as providing four years of support and kind words.

CONTENTS

Understanding population, genetic and antigenic diversity of the poultry red mite (*Dermanyssus gallinae*) to aid development of novel methods for control 1

Student Declaration Form 2

Abstract 3

Acknowledgements 4

List of Figures 16

List of Tables 23

1 General Introduction 29

1.1 The poultry industry 30

1.2 Poultry housing systems 30

1.2.1 Cage system 30

1.2.2 Barn system 30

1.2.3 Free-range system 31

1.2.4 Organic 31

1.3 Poultry parasites 31

1.3.1 Mites (Arthropoda: Chelicerata: Arachnida: Acari) 32

1.4 The poultry red mite (*Dermanyssus gallinae*) 32

1.4.1 Taxonomy 33

1.4.2 Life cycle of *D. gallinae* 33

1.4.3 Clinical signs 34

1.4.4 Host specificity 35

1.4.5 Microbiome of *D. gallinae* 36

1.4.6 *Dermanyssus gallinae* as a vector for disease 36

1.4.7 Impact of *D. gallinae* 37

1.4.8 Host location 39

1.4.9 Pheromone cues in *D. gallinae* 40

1.4.10 Why is poultry red mite an increasing issue? 40

1.4.11 Current control of *D. gallinae* 41

1.5 Vaccination 49

1.5.2 Natural infestation of *D. gallinae* and hen immune response 51

1.5.3 Vaccination against *D. gallinae* 52

1.5.4 Summary of current vaccine research against *D. gallinae* 54

1.5.5 Considerations for vaccinating against *D. gallinae* 55

1.6 Summary of control methods 56

1.7 Studying genetic diversity 56

1.7.1 Molecular markers 56

1.7.2 Utilising bioinformatics 57

1.8 Genetics of *D. gallinae* 57

1.8.1 Genetic diversity of *D. gallinae* 58

1.8.2 Internal transcribed spacer regions 58

1.8.3 Cytochrome c oxidase subunit I (COI) 59

1.9 Haplodiploidy in Acari species 60

1.9.1 Sexual selection under haplodiploidy 60

1.9.2 Evolutionary genetics of haplodiploidy 60

1.9.3 Selective advantages of arrhenotoky vs pseudo-arrhenotoky 61

1.9.4 Haplodiploidy in *D. gallinae* 62

1.10 Concluding remarks 62

1.11 Aims, objectives and hypotheses 63

1.11.1 Objective 1: Sample of *D. gallinae* across the United Kingdom and Europe 63

1.11.2 Objective 2: Identify and validate genetic markers for *D. gallinae* 63

1.11.3 Objective 3: Conduct genome-wide genetic analysis to assess *D. gallinae* population structure and regional variation 63

1.11.4 Objective 4: Studying the occurrence of genotypes associated with acaricide resistance to pyrethroids 63

1.11.5 Objective 5: Assessment of antigenic diversity at loci encoding anti-*D. gallinae* vaccine candidates 63

2 General Methodology 64

2.1 Mite collection 65

2.1.1 Mite collection during visits to UK farms 65

2.1.2 Remote mite collection from UK farms 66

2.1.3 Remote mite collection from European farms 68

2.2 Handling *D. gallinae* in the laboratory 68

2.3 Routine DNA Extraction from whole mites 68
2.4 RNAse treatment 68
2.5 DNA quantification systems 69
2.5.1 Nanodrop 69
2.5.2 Qubit 69
2.6 Primer design and use 69
2.7 Polymerase chain reaction (PCR) 70
2.8 Tris-borate-ethylenediaminetetraacetic acid (TBE) buffer preparation 70
2.8.1 Stock solution of TBE 70
2.8.2 Working solution of TBE 70
2.9 Agarose gel electrophoresis 70
2.10 PCR purification 71
2.11 Sanger Sequencing 71
3 Sampling *Dermanyssus gallinae* in the United Kingdom and Europe 72
3.1 Introduction 73
3.1.1 Prevalence of *D. gallinae* in the United Kingdom 73
3.1.2 Prevalence of *D. gallinae* in Europe 73
3.1.3 Aims and Hypotheses 74
3.2 Methodology 74
3.2.1 Mite collection 74
3.2.2 Ethical approval 74
3.2.3 Questionnaire 74
3.3 Results 75
3.3.1 Sample distribution across the United Kingdom 75
3.3.2 Temporal collection from UK farms 77
3.3.3 Sample distribution across Europe 77
3.3.4 Epidemiological data collection via questionnaire 81
3.4 Discussion 90
3.4.1 Prevalence of *D. gallinae* across the UK 90
3.4.2 UK farm overview 91
3.4.3 Representation across the United Kingdom 91
3.4.4 Factors associated with *D. gallinae* 92

3.4.5 Conclusion 94

4 Genetic Markers for *Dermanyssus gallinae* 95

4.1 Introduction 96

4.1.1 Genetic markers 96

4.1.2 Mitochondrial DNA 97

4.1.3 Advantages of mtDNA markers 97

4.1.4 Mitochondrial vs nuclear gene substitution rate comparison 98

4.1.5 Studies on *Dermanyssus gallinae* mtDNA 99

4.1.6 Mitochondrial Cytochrome c oxidase subunit 1 (COI) 99

4.2 Aims and hypotheses 100

4.2.1 Aim of the study 100

4.2.2 Hypotheses 100

4.3 Methodology 100

4.3.1 Sample collection and distribution 100

4.3.2 Polymerase chain reaction (PCR), primer design and amplicon sequencing 104

4.3.3 Nucleotide analysis 104

4.3.4 Statistical tests 104

4.3.5 Phlyogenetic analysis 105

4.3.6 Alignment to published GenBank sequences 107

4.4 Results 107

4.4.1 Nucleotide sequence analysis 107

4.4.2 United Kingdom and Greece 116

4.4.3 Nucleotide and haplotype diversity by production system 116

4.4.4 Neutrality tests 118

4.4.5 COI sequence variation in the United Kingdom 123

4.4.6 Intra-farm variation: Greece 128

4.4.7 Phylogenetic analysis of a 565-bp COI fragment in *D. gallinae individuals from* the United Kingdom 129

4.4.8 Phylogenetic analysis of a 565-bp COI fragment in *D. gallinae* individuals from European and UK individuals 130

4.4.9 Network analysis 132
4.4.10 Comparative analysis with sequences published in GenBank 133
4.5 Discussion 135
4.5.1 Geographical clustering 135
4.5.2 Genetic diversity of COI in UK *D. gallinae* populations 136
4.5.3 Intra-farm genetic variation 138
4.5.4 Neutrality testing 139
4.6 Conclusions 140
5 Genome-wide genetic analysis: SNP Discovery through the GATK pipeline 141
5.1 Introduction 142
5.1.1 Genetic diversity 142
5.1.2 Quantifying genetic diversity 145
5.1.3 Single nucleotide polymorphisms (SNPs) 145
5.2 Aims and hypotheses 149
5.2.1 Main aim 149
5.2.2 Hypothesis 149
5.3 Methodology 149
5.3.1 Current resources for *Dermanyssus gallinae* 149
5.3.2 The Genome Analysis Toolkit 150
5.3.3 Phase one 151
5.3.4 Phase two 155
5.3.5 VCF Intersection 156
5.3.6 Second round of SNP calling 156
5.3.7 Summary of pipeline process 157
5.4 Results 159
5.4.1 BWA-MEM mapping 159
5.4.2 SNP discovery round 1 159
5.4.3 Intersected VCF files: Hannover read sets one and two 160
5.4.4 SNP discovery round 2 160
5.4.5 Intersected VCF files: Hannover read sets one and two Round 2 161

5.4.6 Substitution type 161

5.4.7 Co-variate analysis: Comparison of RMSE values between rounds 163

5.4.8 Read sets comparisons through each SNP calling round 171

5.5 Discussion 172

5.5.1 SNP database for future studies 172

5.5.2 Biological validation 172

5.5.3 GATK pipeline complications and solutions for *D. gallinae* 173

5.5.4 SNP identification in related species 174

5.5.5 Future considerations 178

5.5.6 Conclusions 179

6 Genome-wide genetic analysis: SNP genotyping of *D. gallinae* using an NGS multiplex platform 180

6.1 Population genetics 181

6.1.1 Population genomics 181

6.1.2 Episodes of selection 181

6.1.3 Population Structure and Phylogeography 182

6.1.4 Estimating allele frequencies 182

6.1.5 SNP genotyping 183

6.2 Aims and hypotheses 185

6.2.1 Aim of the study 185

6.2.2 Hypotheses 185

6.3 Methods 186

6.3.1 Sample selection 186

6.3.2 Geographical clustering 188

6.3.3 Selection of high-quality SNPs for assay development 188

6.3.4 Annotation of target SNPs 191

6.3.5 Validation of SNPs through PCR and Sanger sequencing 191

6.3.6 Eurofins Mid-Plex Genotyping method 194

6.3.7 Data processing 195

6.3.8 Population genetic analyses 195

6.4 Results ... 196
6.4.1 Assay optimisation ... 196
6.4.2 Marker characteristics for the primary SNPs selected for Mid-Plex analysis ... 198
6.4.3 Marker characteristics for the secondary SNPs selected for Mid-Plex analysis 202
6.4.4 Biological validation: PCR results ... 202
6.4.5 Summary statistics ... 202
6.4.6 Nucleotide analysis ... 204
6.4.7 Linkage disequilibrium ... 204
6.4.8 Network analysis of all genotyped samples ... 206
6.4.9 Network analysis of UK farms ... 208
6.5 Discussion ... 211
6.5.1 SNP annotation ... 211
6.5.2 Biological validation of the GATK pipeline ... 212
6.5.3 Genetic diversity ... 212
6.5.4 Population structure ... 213
6.5.5 Considerations and complications faced in SNP genotyping *D. gallinae* samples 217
6.6 Conclusion ... 219
7 Pyrethroid resistance in *Dermanyssus gallinae* and frequency in UK and European populations ... 220
7.1 General introduction ... 221
7.1.1 Resistance to acaricides ... 221
7.1.2 The development of resistance and mechanisms of resistance ... 222
7.1.3 Use of acaricides to control *D. gallinae* and resistance mechanisms ... 222
7.1.4 Pyrethrins/Pyrethroids for the control against *D. gallinae* and resistance mechanisms ... 223
7.1.5 Implications of acaricide resistance ... 224
7.2 background information ... 224
7.2.1 Identification of target site mutations associated with pyrethroid resistance (completed by collaborators) ... 224
7.2.2 Aims and hypotheses ... 227

7.3 Methodology ... 227
7.3.1 Sample selection ... 227
7.3.2 SNP genotyping of M918, L925 and F1534 mutations ... 229
7.4 Results ... 230
7.4.1 Sanger sequencing of pyrethroid mutations of *D. gallinae* populations from UK and Europe ... 230
7.4.2 Mid-plex genotyping of pyrethroid mutations of *D. gallinae* populations from UK and Europe ... 234
7.5 Discussion ... 241
7.5.1 Resistance in UK populations ... 241
7.5.2 Presence and frequency of pyrethroid resistance mutations across Europe ... 243
7.5.3 Future investigations into acaricide resistance in *D. gallinae* ... 244
7.6 Conclusion ... 245
8 Gene-specific genetic analysis: assessment of existing genetic diversity and signatures of selection at loci encoding anti-PRM vaccine candidates. ... 246
8.1 Introduction ... 247
8.1.1 Vaccination against arthropods ... 247
8.1.1 Antigenic diversity ... 248
8.1.2 Chosen vaccine candidates ... 250
8.2 Aims and hypotheses ... 255
8.2.1 Aim of the study ... 255
8.2.2 Hypotheses ... 256
8.3 Methodology ... 256
8.3.1 Sample selection ... 256
8.3.2 DNA extraction ... 257
8.3.3 RNA Extraction ... 257
8.3.4 TurboDNAse treatment ... 258
8.3.5 Reverse Transcription PCR (RT-PCR) ... 258
8.3.6 Primers ... 259
8.3.7 PCR and agarose gel electrophoresis ... 262
8.3.8 PCR purification and Sanger sequencing ... 262

8.3.9 Nucleotide analysis 262
8.3.10 Neutrality tests 262
8.4 Results 263
8.4.1 Cathepsin D cDNA 263
8.4.2 Vitellogenin 265
8.4.1 SNP genotyping of vaccine candidates using the Mid Plex sequencing assay .. 273
8.5 Discussion 276
8.5.1 Vitellogenin 276
8.5.2 Cathepsin D 279
8.5.3 Paramyosin and Tropomyosin 280
8.5.4 Complications with amplification via PCR 280
8.6 Conclusion 281
9 General Discussion 282
9.1 Population structure and genetic diversity of D. gallinae 284
9.1.1 Population structure and genetic diversity across production systems 286
9.1.2 Temporal changes in genetic diversity in *D. gallinae* populations 286
9.2 Future work 287
9.3 Contribution to the field of research 288
References 289
10 Supplementary 354
10.1 Mite sampling pack contents 354
10.1.1 Questionnaire regarding *D. gallinae* in the United Kingdom 354
10.1.2 Instructions for setting traps for *D. gallinae* 360
10.1.3 Information sheet provided to UK farmers 362
10.2 Statistical output regarding nucleotide and haplotype diversity for COI fragments 363
10.2.1 Nucleotide diversity Dunnett's multiple comparison test 363
10.2.2 Haplotype diversity Tukeys multiple comparison test 367
10.2.3 Haplotype and nucleotide diversity comparison of Greece and United Kingdom 371
10.2.4 Nucleotide and haplotype diversity comparison between UK and European countries 372
10.2.5 Nucleotide and haplotype diversity comparison by geographical clustering ... 374

10.2.6 Nucleotide and haplotype diversity comparison of different UK production systems 375

10.2.7 Neutrality test: Sliding window comparison for the full dataset 376

10.2.8 Neutrality test: Sliding window comparison for Greece 377

LIST OF FIGURES

Figure 1: The life cycle of the chicken mite, *Dermanyssus gallinae* (De Geer, 1778). Illustration from Sparagano et al. 2014.(34) 34

Figure 2: Graph depicting changes in egg production via differing production systems between 2006-2019 (100). 41

Figure 3: Current status of arthropod ectoparasite vaccines, reproduced from (186). The graph depicts the number of publications and/or patents which refer to vaccines using ectoparasite-derived antigens over the total number of publications and/or patents which appeared between 2017- 2019 50

Figure 4: An illustration of the set up used for storing poultry red mite after collection in T-75 flasks in the laboratory 65

Figure 5: Summary of sampling advice provided to farmers in four steps: 1) Receiving the mite sampling pack 2) Placing the traps around the poultry houses 3) leaving the traps for 7-14 days and 4) packaging traps into Ziplock bags to send to the RVC 67

Figure 6: Approximate location of farms sampled for *D. gallinae* across the United Kingdom with 15 farms in England, four in Scotland, four in Northern Ireland and one in Wales 75

Figure 7: Map showing the origin of all *D. gallinae* populations sampled from 16 European countries 78

Figure 8: Map showing the origin of all *D. gallinae* populations analysed in the study, spread across 14 European countries. Six geographical clusters are highlighted in the key (red, orange, blue, green, grey and dark grey pointers). Pointers correspond to approximate locations of closest town or regions for each country outlined in Tables 9 and 10. Geographical clusters are used for some genetic analyses (see below) 101

Figure 9: Nucleotide diversity (Pi) for all samples and individual countries. Comparison of means between countries statistically analysed using a one-way ANOVA. Bars indicated with different letters (ABCD) were found to be significantly different (P<0.05). 109

Figure 10: Haplotype diversity for all samples and individual countries. Statistically analysed using an Ordinary one-way ANOVA with a Dunnett's comparison post-hoc test ran on GraphPad Prism 8.4.3. Standard error bars included Bars indicated with different letters (ABC) were found to be significantly different (P<0.05) 110

Figure 11: Comparison of nucleotide diversity in all samples and individual European countries with the UK in a 565-bp fragment COI gene for *D. gallinae*. Statistically analysed using a one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Standard error bars included. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001. 111

Figure 12: Comparison of haplotype diversity in all samples and individual European countries to the UK in a 565-bp fragment COI gene for *D. gallinae*. Statistically analysed using an Ordinary one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Standard error bars included. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001. 112

Figure 13: Difference in nucleotide diversity (Pi) in a 565-bp fragment of the COI gene for *D. gallinae* from six geographical clusters in comparison to the full data set. Standard error bars shown. Statistically analysed using a y one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001. 115

Figure 14: Difference in mean haplotype diversity in a 565-bp fragment of the COI gene for *D. gallinae* from six geographical clusters in comparison to the mean haplotype diversity for the full data. Standard error bars shown. Statistically analysed using a one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001 or a (★) representing an adjusted P-value of P<0.05. 115

Figure 15: Graphs depicting the difference in (A) nucleotide diversity and (B) haplotype diversity when comparing different UK production systems Standard error bars shown. Statistically analysed using an Ordinary one-way ANOVA with a Tukey's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Statistical significance of P<0.0001 indicated by (★) and P<0.05 by a (★). 117

***Figure 16**: Sliding window comparison of Fu and Li's D and F tests with Tajima's D and nucleotide diversity for the full dataset. Sliding window computed on DnaSP from site 1 to 564, with a sliding window length of 100 sites and step size 25. Significance for Fu and Li's D and F: #, P<0.10; *, P<0.05; **, P<0.02 and for Tajima's D: # P<0.10; ★ P<0.05; ★★P<0.01; ★★★ P<0.001.* 121

Figure 17: Sliding window comparison of Fu and Li's D and F test with Tajima's D and nucleotide diversity for Greece. Sliding window computed on DnaSP from site 1 to 564, with a sliding window length of 100 sites and step size 25. Significance for Fu and Li's D and F: #, P<0.10; ★, P<0.05; ★★, P<0.02 and for Tajima's D: # P<0.10; ★ P<0.05; ★★ P<0.01; ★★★ P<0.001 ... 122

Figure 18: Graph depicting the number of variable sites from a 565-bp fragment of the COI gene in *D. gallinae* isolates demonstrating intra-farm variation attributed to UK production systems (free-range (circle), intensive (square), or both(triangle)) 127

Figure 19: Phylogenetic tree of partial COI sequences representing *D. gallinae* isolates collected in the UK, inferred using the Tamura 3-parameter and maximum-likelihood (358). A discrete Gamma distribution was utilised to model evolutionary differences among sites (5 Categories

(+G, parameter = 0.0500)). A total of 559 positions were used in the analysis, encoding 39 nucleotide sequences. All evolutionary analysis was completed with MEGA X (353). Countries from the UK are indicated as follows: England = no colour, Red = Wales, blue = Scotland, green = Northern Ireland. To differentiate production systems, intensive farms are indicated by a black star. 129

Figure 20: (A) Phylogenetic tree of all European and UK isolates sequenced as part of this study. Inferred using the Tamura 3-parameter and maximum-likelihood with 1000 replicates (Tamura, 1992). A gamma distribution was utilised to model evolutionary differences (shape parameter = 0.5). A total of 565 positions were used in the analysis, encoding 196 nucleotide sequences. All evolutionary analysis was completed with MEGA X (353).(B) Bayesian phylogenetic tree of all European and UK isolates sequenced as part of this study. Inferred using the HKY+G+I model with 2 runs, 5,000,000 generations and 25% Burnin. A total of 565 positions were used in the analysis, encoding 196 nucleotide sequences. Evolutionary analysis completed on TOPALi (354) and edited on iTOL (355). 131

Figure 21: Network analysis of all European and UK *D. gallinae* isolates sequenced in the study with the three main haplogroups labelled, A, B and C. Median-joining tree analysis was completed on Network 5.0.0.3. Countries are colour coded in the key in the bottom right. A total of 554 positions were used in this analysis, encoding 195 nucleotide sequences. 132

Figure 22: Network analysis of *D. gallinae* isolates sequenced from the UK and Genbank sequences available from Japan (238). Countries are colour coded in the key in the bottom right. Median-joining tree analysis was completed on Network 5.0.0.3. A total of 554 positions were used in this analysis, encoding 139 nucleotide sequences... 133

Figure 23: Network analysis of all European and UK *D. gallinae* isolates sequenced in the study and Japanese sequences available from Genbank (238). Median-joining tree analysis was completed on Network 5.0.0.3. The three main haplogroups are labelled A, B and C. A total of 554 positions were used in this analysis, encoding 270 nucleotide sequences. Colour coded key provided for country identification.. 134

Figure 24: Overview of determinants of genetic diversity**.** Effective population size, mutation rate and linked selection are the main factors affecting diversity. These factors are governed by several other parameters. The direction of the correlation is indicated by the + and – symbols. Selfing – self-fertilization. Adapted from (363) 144

Figure 25: Flowchart of the general approach to single nucleotide polymorphism (SNP) mining from DNA sequence data. STSs: sequence-tagged sites, ESTs: Expressed sequence tags, RE-fragments: Restriction fragments. Reproduced from (452). 148

Figure 26: GATK best practices for germline short variant discovery (SNP and Indels). Sourced from (467). 150

Figure 27: An overview of the process used for SNP identification in *D. gallinae* using GATK. The four key stages involved are outlined 1) Initial round of SNP identification 2) Consolidation of VCF files 3) A second round of SNP calling and 4) Production of final SNP tables 154

Figure 28: Overview of the adapted GATK pipeline used for self-validation for SNP identification in *D. gallinae* Round one and two followed the same phase one, with the exception of yellow stars (*) indicating where reference SNPS were provided in round two. Phase two indicating differences in round one (grey) and round two (teal). 158

Figure 29: Graph showing the number of SNPs for each substitution type in the VCF intersect file of *D. gallinae* transcriptomic read sets compared to the genome assembly. Black; SNPs from A-G/T/C, Dark blue; SNPs from C- G/T/A, Medium blue; SNPs from G- A/T/C, Light blue; SNPs from T- A/C/G 162

Figure 30: Comparison of RMSE values between rounds of SNP discovery in *D. gallinae* for Co-variate analysis. CycleCovariate results for Hannover read set one; top row round 1 SNP calling, left before recalibration and right after recalibration, bottom row for second round of SNP calling, left before recalibration, right after recalibration. Similar results were observed for read set two so were not shown in duplicate.. 163

Figure 31: Comparison of RMSE values between rounds of SNP discovery in *D. gallinae* for the Dinucleotide covariate. DinucleotideCovariate results for Hannover read set one; top row round 1 SNP calling, left before recalibration and right after recalibration, bottom row for second round of SNP calling, left before recalibration, right after recalibration. Similar results were observed for read set two so were not shown in duplicate. 164

Figure 32: Comparison of RMSE values between rounds of SNP discovery in *D. gallinae* for the quality score covariate. QualityScoreCovariate results for Hannover read set one; top row round 1 SNP calling, left before recalibration and right after recalibration, bottom row for second round of SNP calling, left before recalibration, right after recalibration. Pink plots refer to uncalibrated data and blue plots refer to calibrated data. Similar results were observed for read set two so were not shown in duplicate. 165

Figure 33: Gaussian mixture model reports for HRun (1), MQRankSum (2), QD (3) and ReadPosRankSum (4) plotted against FS as part of the VSQR process. Produced on usegalaxy. Probability density function with green areas indicative of high quality and red areas indicative of lower quality displayed in panels 1-4 A. Retained SNPs (black) and filtered SNPs (red) shown in panels 1-4 B. Distribution of the SNPs used to train the model, with green representing SNPs which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being retained shown in panels 1-4 C. SNPs by status (novel or known) shown in panels 1-4 D. 167

Figure 34: Gaussian mixture model reports for MQRankSum, QD and ReadPosRankSum plotted against HRun as part of the VSQR process. Produced on usegalaxy. Probability density function with green areas indicative of high quality and red areas indicative of lower quality displayed in panels 1-3 A. Retained SNPs (black) and filtered SNPs (red) shown in panels 1-3 B. Distribution of the SNPs used to train the model, with green representing SNPs which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being retained shown in panels 1-3 C. SNPs by status (novel or known) shown in panels 1-3 D. 169

Figure 35: Gaussian mixture model reports for ReadPosRankSum and QD plotted against MQRankSum, and ReadPosRankSum plotted against QD as part of the VSQR process. Produced on usegalaxy. Probability density function with green areas indicative of high quality and red areas indicative of lower quality displayed in panels 1-4 A. Retained SNPs (black) and filtered SNPs (red) shown in panels 1-4 B. Distribution of the SNPs used to train the model, with green representing SNPs which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being retained shown in panels 1-4 C. SNPs by status (novel or known) shown in panels 1-4 D. 170

Figure 36: Comparison of genome coverage and average Phred quality score for individual read sets and intersected VCF files for *D. gallinae* SNP identification using GATK. Left: Genome coverage (calculated by % of contigs from genome assembly) for read sets and intersected files for each round of SNP calling, Right: Average Phred quality score between SNP calling rounds. Round one indicated by black and round two indicated by blue colouration......................... 171

Figure 37: SNP filtration process of 32,599 SNPs identified in *D. gallinae* for selection of 92 target SNPs for Eurofins Mid-Plex genotyping assay. Dashed green line indicating change in plan from Sequenom specifications to Mid-plex specifications.. 189

Figure 38: Left: flowchart outlining the process of identifying SNPs in vaccine candidates for incorporation into the SNP panel Right: flowchart outlining identifying genome co-ordinates for acaricide resistance markers (KDR1 and KDR3 where KDR is kinase domain receptor. CLC: CLC Main Workbench 20.0.4 (QIAGEN) 190

Figure 39: Coverage classes achieved by proprietary primer mixes for individual samples tested on pooled *D. gallinae* samples for 100 primary SNPs by Eurofins Genomics during assay optimisation. Sample names have been removed for confidentiality, with one sample present per column, separation indicated by black lines (I). 197

Figure 40: Total coverage classes achieved by proprietary primer mixes tested on pooled *D. gallinae* samples for 100 primary SNPs by Eurofins Genomics during assay optimisation 197

Figure 41: Frequency of substitution changes. Top: Frequency for entire dataset Bottom: Frequency of target 100 SNPs.. 198

Figure 42: Annotation associated with the primary SNPs identified in *D. gallinae* and selected for SNP genotyping including exon, intron or not identified 199

Figure 43: Graph depicting quality score and read depth scores for SNPs with corresponding .gff annotation vs. SNPs with no annotation found. Red dots representing SNPs with no associated annotation and grey representing SNPs associated with an annotation in the .gff file 201

Figure 44: Network analysis of pooled *D. gallinae* samples from the UK and the rest of Europe genotyped by NGS multiplex sequencing (238). Countries are colour coded in the key. Median-joining tree analysis was completed on Network 10.0.0.0. A total of 145 positions were used in this analysis, encoding 75 nucleotide sequences 207

Figure 45: Network analysis of pooled *D. gallinae* samples from the UK genotyped by NGS multiplex sequencing by farm (left) and country (right) (238). Individual farms and countries are colour coded in the keys provided. Farms UK6, UK7 and UK 11were sampled on multiple occasions, permitting analysis of temporal variation, where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated. Multiple barns sampled for UK6 are indicated by letters (a) barn 1 (b) barn 2 and (c) barn 3. Median-joining tree analysis was completed on Network 10.0.0.0. A total of 145 positions were used in this analysis, encoding 28 nucleotide sequences 209

Figure 46: Network analysis of pooled *D. gallinae* samples from the UK genotyped by NGS multiplex sequencing BY production system (right)(238). Production systems are colour coded in the key. Farms UK6, UK7 and UK 11were sampled on multiple occasions, permitting analysis of temporal variation, where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated. Multiple barns sampled for UK6 are indicated by letters (a) barn 1 (b) barn 2 and (c) barn 3. Median-joining tree analysis was completed on Network 10.0.0.0. A total of 145 positions were used in this analysis, encoding 29 nucleotide sequences. 210

Figure 47: Schematic diagram of domain II and III of the VGSC. Mutations found in *D. gallinae* are indicated with orange circles, while those found in other species are indicated with black circles. Those mutations that were functionally characterized in Xenopus are framed in a box (646). Figure reproduced from (589). 226

Figure 48: Map showing the origin of 53 *D. gallinae* populations analysed in this study, spread across 15 European countries, where GBR: United Kingdom, DNK: Denmark, FRA: France, PRT: Portugal, ITA: Italy, ALB: Albania, TUR: Turley, ROU: Romania, HRV: Croatia, CZE: Czech Republic, GER: Germany, BEL: Belgium, NL: the Netherlands and GRC: Greece..................................... 227

Figure 49: Nucleotide and amino acid sequence of KDR1 and KDR3 from *D. gallinae* with pyrethroid mutations annotated M918, L925 and F1354 codon positions. 229

Figure 50: Schematic illustration of the distribution of pyrethroid resistance mutations in *D. gallinae* populations across Europe. The classification of mutations was based on visual inspection of sequencing chromatographs and comprised three categories: 'absent', 'present' and 'fixed' (when no background signal was detected at the investigated position). 233

Figure 51: Schematic illustration of the distribution of pyrethroid resistance mutations in *D. gallinae* populations across Europe from SNP genotyping. The classification of mutations was based on presence or absence of multiple alleles comprised to make three categories: 'absent (reference only allele), 'present in 1 population' (heterozygous reference and alternative alleles or heterozygous alternative alleles) and 'fixed in 1 population' (alternative only alleles present). 237

Figure 52: Schematic illustration of the distribution of pyrethroid resistance mutations in *D. gallinae* populations across the UK from SNP genotyping. The classification of mutations was based on presence or absence of multiple alleles comprised to make three categories: 'absent (reference only allele), 'present in 1 population' (heterozygous reference and alternative alleles or heterozygous alternative alleles) and 'fixed in 1 population' (alternative only alleles present). 240

Figure 53: Sliding window analysis of nucleotide diversity across vitellogenin for *D. gallinae.* A sliding window size of 100 was used with a step size of 25. Dashed lines indicating the end and start of consecutive primer pairs. Primer pairs put in chronological order (2,3,4,6,7 and 8) .. 269

Figure 54: Putative conserved domains in *D. gallinae* amino acid sequence as predicted by BLASTP (755, 756) and the CDD database and SPARCLE with interval locations 270

Figure 55: Sliding window comparison of Fu and Li's F and D test and Tajima's D test for all primer pairs across vitellogenin for all *D. gallinae*. Sliding window computed on DnaSP, with a sliding window length of 100 sites and step size 25. Dashed lines indicating where primer pair alignments finish and begin, in chronological order (2, 3, 4, 6, 7 and 8)................................ 272

LIST OF TABLES

Table 1: Viral and bacterial pathogens 'associated' with Dermanyssus gallinae, Table based on information originally published by Valiente Moro et al., updated with information from Huong et al. (48) and edited from George et al. (65) 37
Table 2: Summary of antigens tested as vaccine candidates to control infestation by D. gallinae. * The effects are significant. Abbreviations: M = Mortality, O = Oviposition, up arrow = increase, down arrow = decrease, IEX= Ion-exchange chromatography. Sourced from(204). .. 55
Table 3: Method of approach for sampling of D. gallinae and the number of farms collected through each approach 74
Table 4: Location of farms sampled in the UK including the farm number, county it resides in, the style of production system used, whether D. gallinae isolates were successfully received and if questionnaire data was complete. * indicates a free-range backyard flock. N/A = not available. 76
Table 5: Temporal sampling of three farms (UK11, UK 7 and UK6), county farm resides in, production type and dates of sampling provided.. 77
Table 6: Sample locations from Europe (outside of the UK), including the country and number of individual mites sampled 80
Table 7: Summary of husbandry-related questionnaire data from 18 UK farms sampled for D. gallinae in relation to production system, housing, bedding, chickens and feeding, with the number of farms and % of total farms sampled provided. Questions permitted the respondent to answer with multiple categories causing discrepancies with total % not equalling 100%. ... 83
Table 8: Summary of questionnaire data defining opinions from 18 UK farms sampled for D. gallinae in relation to D. gallinae specific topics including current control measures, with the number of farms and % of total farms sampled provided. Questions permitted the respondent to answer with multiple categories causing discrepancies with total % not equalling 100%. Italics indicating number of farmers reporting specific locations of visible mites 86
Table 9: Location of farms sampled for D. gallinae in the UK including the number of individuals, style of production and geographic cluster 102
Table 10: Sample locations from Europe (outside of the UK) for D. gallinae collection, including the country and number of individual mites sampled 103
Table 11: Nucleotide diversity, average number of nucleotide differences and haplotype diversity for a 565-bp fragment of the COI gene in D. gallinae individuals from individual countries and the full dataset. Generated using DnaSP version 6. 108

Table 12: Nucleotide and haplotype diversity and the average number of nucleotide differences based on six geographical clusters for a 565-bp fragment COI gene in D. gallinae individuals. Generated on DnaSP version 6. 114

Table 13: Nucleotide and haplotype diversity and the average number of nucleotide differences based on production system for a 565-bp fragment COI gene in D. gallinae individuals. Generated on DnaSP version 6. 116

Table 14: Results of Tajima's D test for the whole dataset and each individual country with associated simulated P value based on a 565-bp fragment COI gene in D. gallinae individuals. Data generated on DnaSP version 6. 118

Table 15: Fu and Li's D and F test results with associated P-values for the full dataset, UK (including separate production systems) and individual European countries based on a 565-bp fragment COI gene in D. gallinae individuals. Data generated on DnaSP version 6. 120

Table 16: Variable positions for a 565-bp fragment of the COI gene for UK D. gallinae individuals when compared to the consensus sequence. Information regarding the base pair at the consensus and the mutation(s) present with the number of individual provided. Mutations which were found in a single country from the UK are indicated in the far-right column, with any mutations present in multiple indicated by a dash (-). A * is used to indicate that all individuals belong to a single farm. 124

Table 17: Intra-farm variation observed in D. gallinae individuals for a 565-bp fragment of the COI gene from UK farms. Base pair position, farm, country and information regarding number of samples with the mutation compared to the consensus sequence.................................... 125

Table 18: Intra-farm variation observed in D. gallinae individuals for a 565-bp fragment of the COI gene for different UK production systems. Base pair position relating to mutations detailed in Table 17 with columns indicating presence in free-range farms, intensive farms or in both production systems. 126

Table 19: Base-pair positions from a 565bp fragment of the COI gene from D. gallinae individuals collected from Greek farms displaying Intra-farm variation, including information regarding the number of individuals and percentage of individuals with the mutation.......... 128

Table 20: Input genomic and transcriptomic D. gallinae data utilised in the SNP discovery pipeline. Number of read sets, file size, read type and sequencing platform, and whether quality scores were available for each dataset, are indicated.. 151

Table 21: Summary of annotations applied in the UnifiedGenotyper step of the GATK pipeline for SNP identification in D. gallinae. The summaries of the following variants are outlined: FIsherStrand, MappingQualityRankSumTest, QualbyDepth and ReadPosRankSumTest. 156

Table 22: BWA-MEM mapping results for each transcriptomic read set to the D. gallinae genome assembly with percentage of raw reads according to 5.3.3.1 successfully mapped outlined for each round of SNP discovery 159

Table 23: Results from the first round of SNP calling using the GATK pipeline for identification of SNPs in D. gallinae. File size, sequencing platform, read type, mapping results and number of SNPs identified are labelled for both read sets from the Hannover transcriptomic dataset ... 159

Table 24: *Results from the second round of SNP calling using the GATK pipeline to identify SNPs in D. gallinae. Information is provided for each transcriptomic read set for the number of SNPs identified from each round of SNP discovery. Columns outlining the number of SNPs which passed the filtering process and the number excluded are included for round two of SNP discovery, with the total number of SNPs from the intersect file shown and the SNPs which passed filtration.* 160

Table 25: Type, total number and percentage of substitution types in the final VCF intersect file, produced from intersection of the second round of SNP calling comparing D. gallinae transcriptomic read sets to the genome assembly........ 161

Table 26: Sample locations from UK, including production type, organic status, farm no. and time points where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated 186

Table 27: Sample locations from Europe (outside of the UK), including the country and production type 187

Table 28: Geographical clustering of countries sampled for D. gallinae with assigned cluster number and countries included per country detailed........ 188

Table 29: Primer sequences, length, melting temperature, GC%, presence of secondary structure or primer dimer in primers used to validate ten SNP markers within D. gallinae genomes........ 193

Table 30: Information regarding associated proteins identified in the .gff file, including chromosome number, position relative to contig, reference and alternative alleles, source, feature type (intron or exon), start and end of contig, score, strand (positive or negative), frame and comment on similarity to other known organisms 200

Table 31: Summary of PCR results for biological validation of ten SNP markers from pooled D. gallinae samples from four farms, including amplification success and presence of reference and alternative alleles........ 202

Table 32: Summary genotyping results of 157 SNPs for 108 pooled D. gallinae samples from across Europe and UK with percentage call rates achieved for samples and SNPs 203

Table 33: Summary of 145 SNPs for 75 pooled D. gallinae samples retained for analysis after quality control, including, alternative or heterozygotes called and the quartile ranges for

reference and alternative NA alleles. Discrepancies in the quartile ranges due to uncalled markers (i.e. N's) 203

Table 34: Linkage analysis of 145 SNP markers for 75 pooled D. gallinae samples using LIAN (594). Results shown for each dataset analysed including the full dataset, production systems across Europe, four geographic clusters, UK production system and UK organic status. Number of samples and number of haplotypes included for each dataset provided *Four farms were removed from production type analysis due to information regarding production system utilised not avaliable. ** Total of 25 farms included in organic status analysis, as intensive farms due to inability to be organic. H : mean genetic diversity, V_D: observed mismatch variance, V_e: expected mismatch variance, I_S^A: Standardized index of association, P: calculated significance, L: Simulated 5% critical value 205

Table 35: Linkage disequilibrium across 140 markers genotyped for 75 D. gallinae populations calculated on DNAsp, with all significant sites and Bonferroni corrected sites average D' and number of associated markers 206

Table 36: Nucleotide triplets present in the sequenced VGSC gene segments of D. gallinae mites. The nucleotide triplets and amino acids identified in the GRC1 and GRC2 strains that were tested with bioassays are underlined. Nucleotide triplets/amino acids found in the SUSC strain are indicated in normal font, while those found in the screened populations are indicated in bold font. * indicating Musca domestica numbering. Table reproduced from (592). 225

Table 37: UK farms sampled for the analysis including county, farm no. allocated in PhD thesis, farm name allocated for publication and the production system utilised by the farm. 228

Table 38: Amino acid substitutions in two VGSC domains (II and III) of European D. gallinae populations. The susceptible alleles are indicated with bold font. A forward slash, separating amino acids, indicates that the allele is not fixed in a certain D. gallinae populations. Table reproduced from (592)with names corresponding to paper publication. 232

Table 39: Summary of resistant and susceptible VGCS (domain II and III) allele occurrence in D. gallinae sampled from the UK and the rest of Europe. 235

Table 40: Amino acid substitutions in two VGSC domains (II and III) of European D. gallinae populations. The susceptible alleles are indicated with bold font. A forward slash, separating amino acids, indicates that the allele is not fixed in a certain D. gallinae population. An – indicating samples where data was unavailable to call. 236

Table 41: Amino acid in two VGSC domains (II and III) of UK D. gallinae populations. The susceptible alleles are indicated with bold font. A forward slash, separating amino acids, indicates that the allele is not fixed in a certain D. gallinae. An – indicating samples where data was unavailable to call. Letters a-c indicating barns relating to the same farm. Farms UK6, UK7

and UK 11were sampled on multiple occasions, permitting analysis of temporal variation, where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated. 239

Table 42: Type of extractions conducted with associated sample and production systems used for amplification of vaccine candidates relating to D. gallinae........ 257

Table 43: Primer sequences, length, melting temperature, GC%, presence of secondary structure or primer dimer and corresponding region of mRNA sequence for D. gallinae Cathepsin D 260

Table 44: Primer sequences, length, melting temperature, GC%, presence of secondary structure or primer dimer and corresponding region of mRNA for D. gallinae vitellogenin.... 261

Table 45: Polymorphism identified across Cathepsin-D from D. gallinae cDNA, consensus residue, reference (Cathepsin D mRNA sequence) residue, mutation, no. of populations with reference residue and no. of individuals with SNP outlined 263

Table 46: Summary of nucleotide diversity, haplotype diversity and neutrality tests for cDNA amplified for D. gallinae Cathepsin D 264

Table 47: Nucleotide polymorphisms found in gDNA from four pooled mite samples (ROM6, BEL5, FRA3 and NET7) in comparison to the reference D. gallinae CAT-D sequence........ 264

Table 48: Fragment number, corresponding primers used, resulting amplicon size and no. of individual gDNA D. gallinae and pooled D. gallinae extracts for six fragments of vitellogenin amplified for Sanger sequencing 265

Table 49: Polymorphism identified across vitellogenin according to primer pair, consensus reside, reference (vitellogenin mRNA sequence) residue, mutation, no. of individuals with reference residue and no. of individuals with SNP/indel outlined. ^ indicating an insertion between two nucleotide positions. – indicating no nucleotide present at that site and a / indicating two alternative alleles present at tri-allelic sites 267

Table 50: Nucleotide and haplotype diversity of D. gallinae for vitellogenin, primer pair, sequence length, sequence conservation, no. of samples included in analysis, nucleotide diversity, average number of nucleotide differences, haplotype diversity and number of haplotypes........ 268

Table 51: Haplotype diversity for vitellogenin fragments amplified from D. gallinae, with summary of haplotypes identified in one than one sample 270

Table 52: Tajima's D values calculated for each primer pair covering vitellogenin for D. gallinae isolates, corresponding region of vitellogenin covered, number of samples involved in analysis and statistical significance 271

Table 53: Fu and Li's D and F values calculated for sequence generated from each primer pair for D. gallinae isolates for vitellogenin, no. of segregating sites, D and F values provided by

biallelic positions only and D and F statistic with associated P value from DnaSP version 5 also provided 271
Table 54: Mid-plex SNP genotyping results for 17 SNPs across four loci (Cat-D, paramyosin, tropomyosin and vitellogenin) of 76 pooled D. gallinae samples with number of reference, alternative and NA (failed to amplify) samples and the percentage of the total population (defined here as all 76 populations studied). 274
Table 55: Total number of polymorphisms and non-synonymous vs synonymous mutations in Cathepsin D, Paramyosin and Vitellogenin fragments amplified from D. gallinae pooled samples through Sanger sequencing and SNP genotyping 275
Table 56: Dunnett's comparison test for nucleotide diversity on a fragment of the COI gene for D. gallinae isolates 366
Table 57: Tukey's comparison test for haplotype diversity on a fragment of the COI gene for D. gallinae isolates 370
Table 58: Tukey's multiple comparison test comparing nucleotide diversity for UK, Greece and full dataset for fragment of the COI gene for D. gallinae 371
Table 59: Tukey's multiple comparison test comparing haplotype diversity for UK, Greece and full dataset for fragment of the COI gene for D. gallinae 371
Table 60: Sidak's multiple comparison test for comparison of nucleotide diversity of individual countries to the UK for fragment of the COI gene for D. gallinae 372
Table 61: Dunnett's multiple comparison test for comparison of haplotype diversity between the UK and individual countries for fragment of the COI gene for D. gallinae 373
Table 62: Dunnett's multiple comparison test for comparison of nucleotide diversity by geographical clustering for fragment of the COI gene for D. gallinae 374
Table 63: Dunnett's multiple comparison for comparison of haplotype diversity by geographical clustering for fragment of the COI gene for D. gallinae 374
Table 64: Tukey's multiple comparison test for comparison of nucleotide diversity based on production system for fragment of the COI gene for D. gallinae 375
Table 65: Tukey's multiple comparison test for comparison of haplotype diversity by production system for fragment of the COI gene for D. gallinae 375
Table 66: Sliding window output for Fu and Li's D and F test and Tajima's D for the full dataset. Significance for Fu and Li's D and F: #, P<0.10; *, P<0.05; **, P<0.02 and for Tajima's D: # P<0.10; * P<0.05; ** P<0.01; *** P<0.001 376
Table 67: Sliding window output for Fu and Li's D and F test and Tajima's D for Greece. Significance for Fu and Li's D and F: #, P<0.10; *, P<0.05; **, P<0.02 and for Tajima's D: # P<0.10; * P<0.05; ** P<0.01; *** P<0.001 377

1 GENERAL INTRODUCTION

1.1 THE POULTRY INDUSTRY

With the global human population expected to exceed nine billion by the year 2050, sustainable food security is now an important focus (1). Poultry are a highly efficient source of animal-derived protein, including both meat and eggs (2), and have become an influential sector of animal production (3). In 2016 in the United Kingdom, approximately half (49%) of all meat consumed was poultry, equal to almost all beef, lamb and pork consumption combined (4). Across the UK, in 2016, there were over 2,500 poultry farms with the meat industry producing ~875 million chickens, ~17 million turkeys, ~16 million ducks and ~250,000 geese (4). In 2019, the industry estimate for UK laying flock size was 42 million hens (5), with approximately 13.1 billion eggs eaten per annum and 36 million consumed every day (5). Estimates from DEFRA indicate that 42% of eggs come from hens in laying cages, 56% from free-range (including 3% organic) and 2% from barn systems (6) (See Section 1.2).

1.2 POULTRY HOUSING SYSTEMS

The UK chicken egg production industry is split into four major systems:

1.2.1 Cage system

Conventional 'battery' style cages have been banned since 1st January 2012 (EU Council Directive 1999/74/EC) (7) across the EU. In the UK they were replaced by larger 'enriched' colony cages which currently provide 750cm^2, a nest box, perching space and a scratching area to enable birds to exhibit natural behaviours. In most cases, the enriched colony cages have been designed to accommodate 40-80 chickens (8).

1.2.2 Barn system

Hens raised under a barn system are free to roam around a house with a maximum stocking density of nine hens per square meter permitted under the EU welfare of Laying Hens Directive. Similar to the free-range system, perches are required (15cm per hen) and litter must account for one third of the ground surface, to enable hens to exhibit scratching and dust bathing behaviours. Additionally, communal nests, or one nest box per five hens, should be provided, with a nesting area of 10 birds/m^2. Access to feeders and drinkers should be raised from the floor; for linear feeders this should be at least at 10cm per hen, for circular feeders at least 4cm per hen. To give hens an optimum day length throughout the year electric lighting should be provided (8).

1.2.3 Free-range system

EU egg marketing legalisation stipulates that hens must be given continuous daytime access to runs which are predominantly covered by vegetation and a maximum stocking density of 2,500 birds per hectare in order to be classified as 'free range'. The hen house must also comply with the same regulations that barn systems must adhere to, with a maximum stocking density of nine hens per square metre permitted. In addition, hens should be provided with nest boxes, adequate perches (15cm per hen) and litter that should account for one-third of the available floor space (8).

1.2.4 Organic

Hens classified as producing organic eggs are always free range. The main differences between organic and other non-caged egg production systems include factors such as feed, medication and animal welfare. Hens must not be fed animal by-products, the use of chemicals are limited and it is not permitted to use genetically modified crops (9). EU Organic Regulations stipulate the rules for hen house conditions, which include a maximum flock size of 3,000 hens and a stocking density of six hens per square meter. Hens should be provided with perches (18cm per hen), nest boxes and litter should account for one-third of the ground surface, the remainder to be used for scratching and dust bathing (8).

1.3 POULTRY PARASITES

Animals that are raised and utilised for agriculture, including poultry, are hosts for a multitude of parasites and can attract pestiferous insects (10, 11). Within the poultry industry, parasites are ubiquitous across all production systems, from small backyard flocks to huge commercial operations. They result in significant economic losses, impact hen welfare(10) and negatively affect profit by adversely influencing production parameters.

Among the variety of parasites that the poultry industry faces there are a number of major ectoparasites, primarily ticks, fleas, bugs, lice and mites. The type of poultry production system can influence exposure to parasites and the subsequent burden of infection/infestation. Whilst confinement favours parasites with a short life cycle and direct transmission, free-range systems provide greater opportunities for parasites that depend on an intermediate host. It should also be noted that in general parasitism in poultry is a flock level problem as the impact for one individual bird is of relatively small economic importance (10). Generally speaking, detection and identification is important and easier for parasites which reside on birds (e.g. northern fowl

mites, hard ticks, lice or stick tight fleas) in comparison to parasites which reside elsewhere and only use the bird to feed (e.g. soft ticks, poultry red mite or bed bugs) (10).

1.3.1 Mites (Arthropoda: Chelicerata: Arachnida: Acari)

Whilst insects (Arthropoda: Mandibulata: Insecta) are commonly thought of as being the most numerous animals on the earth, it is possible that mites and ticks could be just as diverse and equally as abundant (1-3). Current estimates suggest there could be as many as 1-3 million mite species once identification has been completed. Around the world, mites reside in every continent, including Antarctica, occurring in arboreal, terrestrial, aquatic and parasitic habitats. They are found inhabiting a vast range of environments, from the inside of caves, freshwater streams and lakes, deep within soil, house dust, and stored food products to intertidal zones. Mites can be parasites of reptiles, arthropods, mammals and birds, both ectoparasitic (e.g. poultry red mite infesting poultry) or endoparasitic (e.g. found in the tracheae of honeybees), as well as free-living detrivores or predators in dung, compost, nests, soil litter or rotting wood, or can feed on nectar, fungi, pollen and other plant material. This level of diversity has evolved over time with date estimates produced from mitochondrial genome sequences indicating divergence at least 450 million years ago (Cambrian period) (12).

Mites belong to the subclass Acari (13). Out of the ~40,000 named mite species the vast majority are deemed to be harmless or are providers of useful ecosystem services from a human standpoint, but a small proportion are considered to be pests of livestock, crops, humans and/or animals as well as having a vectoral role in the transmission of some animal and human diseases (12). Currently at least 2,500 known species of mites are closely associated with birds, spanning 40 families and covering all habitats on the body (14). Avian taxa are associated with acarine parasites, even species which lack feathers (15). Broadly speaking, mites that affect birds can be placed into two categories; (1) those that primarily dwell near or in the nest and (2) those that reside mainly on the host's body. The most comprehensively studied nest-dwelling mites are from the genera *Dermanyssus*. Genus *Dermanyssus* Duges 1834 currently consists of 24 hematophagous mite species which primarily parasitise birds (16, 17), although several species remain cryptic (i.e. species that cannot be distinguished from one another based on morphological analysis alone, despite reproductive isolation) (18).

1.4 THE POULTRY RED MITE (*DERMANYSSUS GALLINAE*)

Dermanyssus gallinae (de Geer, 1778) is an obligatory blood feeding ectoparasite (19). Adults typically have a pear-shaped body, covered by grey chitinous armour that is partially translucent. They range in size from 0.7-1.0 mm long and 0.4-0.5 mm wide (20). *Dermanyssus gallinae* has a

wide global distribution with reports from Algeria, Denmark, France, Italy, Japan, Poland, Romania, Serbia, South Korea, the Netherlands and the UK, where up to 90% of layer hen farms can be infested (21-27). Prevalence figures of ~34% (± 0.075%) have been reported in farms from North-East Tunisia (28) and 95.8% (95% CI: 79.8–99.3%) of Portuguese farms (29). Hamidi *et al.* (2011) reported that across Kosovan free-range farms, 50% were affected by *D. gallinae* (30). *Dermanyssus gallinae* spends the majority of its life cycle living separately from the host, residing in cracks and crevices of poultry houses, bird nests and cages (26), entering through wall inlets, air chimneys, open wall fans or via staff, crates, cages or wild birds (30, 31). Poultry red mites most commonly parasitise birds during hours of darkness (19), remaining on the avian host for approximately 30-90 minutes during feeding (21). Typically females and nymphs suck blood most frequently, males feed less frequently and larvae never feed (20).

1.4.1 Taxonomy

By morphology, *D. gallinae* is described as a member of the Phylum Arthropoda, Class Arachnida, Order Acari, Sub-order Mesostigmata and Family Dermanyssidae. Despite the genus *Dermanyssus* harbouring economically important species, the classification of species has been in a state of confusion (16). After reclassification of the original 56 members of the genus *Dermanyssus*, 23 members remain (18).

1.4.2 Life cycle of *D. gallinae*

The life cycle of *D. gallinae* was first described by Wood (1917) and development consists of five stages: egg, larva, protonymph, deutonymph and adults (Figure 1) (20). Adults will mate shortly after moulting and, after fertilization has occurred, females will begin depositing eggs between 0.5 and three days after feeding. They will repeat the process of feeding and depositing eggs multiple times, with females capable of producing approximately 30 eggs in a lifetime, (19). Eggs are small (270 x 400 µm), smooth, oval and have a pearly white colouration. Emergence of larvae can occur in two to three days if conditions are warm (28-30°C). Larvae have six legs and a white colouration and after one day larvae will moult without feeding to the protonymph stage, which possesses eight legs. After feeding, protonymphs moult to develop into deutonymphs, and after another blood meal will become adults (♂ or ♀) (19). Under optimal conditions, i.e. 20-25°C and high relative humidity (>70%), the entire life cycle from egg to egg can be completed in approximately 7-10 days (20, 32).

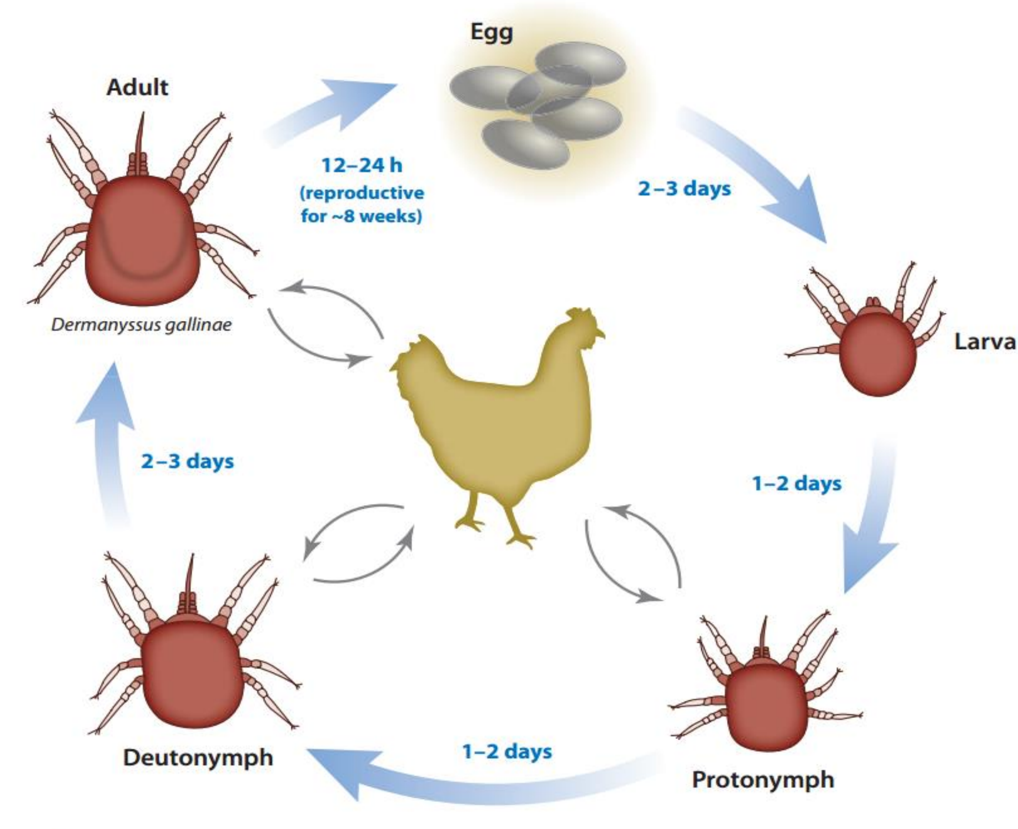


Figure 1: *The life cycle of the chicken mite,* Dermanyssus gallinae *(De Geer, 1778). Illustration from Sparagano* et al*. 2014.(33)*

Nordenfors *et al.* (1999) conducted research into how temperature and humidity affect *D. gallinae* in relation to oviposition, longevity and moulting (34). They noted that mites can survive up to ~nine months in the absence of a suitable host when kept in temperatures of between 5-25°C, but temperatures exceeding 45°C and falling below -20°C were lethal. This means it is feasible that *D. gallinae* infesting one flock is capable of surviving within housing systems during depopulation and cleaning and be able infect the subsequent flock (11).

1.4.3 Clinical signs

In temperate climates, infestations can be seen all year-round with an average density of ~50,000 mites per hen (35). At infestation levels of >150,000 mites per hen significant impacts on hen welfare can be observed, including an increase in restlessness, irritation, cannibalism, feather pecking, anaemia and hen mortality. Typical infestation levels of approximately 50,000 mites per bird can also compromise hen health and welfare (35).

Research has been conducted focusing on the implications of *D. gallinae* infestation on hen welfare. Maurer *et al.* (1993) conducted an experimental infestation study and discovered that

hens left their typical resting and sleeping places on perches to spend the night in the litter area. They hypothesised that this is the result of mite density being lower in the litter than around the perches (36). Wójcik *et al.* (2000) completed a study focusing on *D. gallinae* at three fully automatic egg-laying farms in the Toruń region. They noted an elevated *D. gallinae* occurrence at one of the farms, detecting all the signs of dermanyssosis (restlessness, irritation, cannibalism, feather pecking, anaemia and hen mortality). They also reported decreased body weight, signs of anaemia and exhaustion, and an increase in hen mortality between unaffected (1%) and infested birds (4%) (37). In the non-contaminated farms studied, they found that egg laying rates typically amounted to 91-93%, whilst in farms contaminated with *D. gallinae* production dropped to 80-82%, indicating a drop-in laying rate of 11-13%. A study by Kilpinen *et al.* (2005) demonstrated an increase in self-grooming and social feather pecking activity in infested birds, both of which represent important indicators for welfare. They also noted differences in weight gain between non-infected and infected birds and found that infested birds were generally more active. While this was not statistically significant, they hypothesised that this influence body weight due to increased restlessness during the day and night (35). Arkle *et al.* (2006) found a significant ($P<0.05$) correlation between total mite population size and bird mortality. They suggested that increased mite numbers can actively contribute to increased total bird losses (38).

Additional research has shown that indicators of stress, e.g. corticosterone levels, increased by up to 2.5 times in hens exposed to *D. gallinae* infestation (39). Sokół *et al*. (2008) studied the activity of specific antioxidant enzymes (superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GPx)) in infested and non-infested chickens. Their findings included significant increases in activity for SOD and GPx and decreased activity for CAT, which they suggest could be an indication of oxidative stress occurring as a consequence of long term infestation with *D. gallinae* (40).

1.4.4 Host specificity

Currently, *D. gallinae* has been demonstrated to affect 28 avian host species, including the domestic hen, *Gallus gallus domesticus* (Galliformes). Other common hosts include canaries, domestic doves and wild doves, as well as other birds which are found to have associations with layer farms (18). *Dermanyssus gallinae* has been shown to display plasticity in terms of host specificity; besides birds *D. gallinae* is capable of feeding on mammals, including horses, rodents and humans (41). Work by Roy et al., (42, 43) and Roy and Buronfosse (44) has suggested that *D.* gallinae consists of a species complex, including at least two cryptic species. The spectrum of hosts differs for each of these cryptic species;- *D. gallinae* s. str. has been recorded in poultry

farms and on other bird species, except for pigeons, whilst *D. gallinae* L1 has only been recorded from pigeons (in the USA and France). Work in Sweden showing significant differences in the internal transcribed spacer (ITS) I region of cryptic species supported this, revealing the presence of two distinct genotypes; one from wild host origin and one from domesticated hosts. They conducted genetic analysis of the small subunit (SSU) gene, 5.8S ribosomal RNA (rRNA) and two ITS of the rRNA genes for *D. gallinae* collected from wild bird populations and laying poultry farms. Results demonstrated identical SSU rRNA sequence for all mites but one ITS1 genotype for mites collected from domesticated chickens and a different ITS1 genotype for wild birds. They noted that the nucleotide differences were not all randomly distributed but appeared as compensatory base changes present in the putative paired stems of the RNA transcripts (45).

1.4.5 Microbiome of *D. gallinae*

Lima-Barbero et al. (2019) used a metaproteomics approach to characterise the alphaproteobacteria in the microbiota of *D. gallinae*. The analysis resulted in a total of 2837 peptide assignments, identifying 11 bacterial proteins representing 53 genera of alphaproteobacteria. The top five genera represented were *Sphingomonas*, *Bradyrhizobium*, *Rhodopseudomonas*, *Methylobacterium* and *Wolbachia* spp. When looking at taxonomic relative abundance, 74% had previously been identified by association as environmental bacteria, 20% identified in multiple mite species (including *D. gallinae*) and 44% have been discovered in arthropod species other than mites, and in humans. They identified potential pathogenic bacteria of the genera *Roseomonas*, *Inquilinus* and *Neorickettsia.* In addition, they demonstrated differential alphaproteobacterial microbiota composition for different developmental stages and suggested that the bacteria could have functional implications in relation to metabolic pathways linked to blood feeding (46).

1.4.6 *Dermanyssus gallinae* as a vector for disease

Dermanyssus gallinae has been suggested to play a role in the transmission of multiple pathogenic agents that are capable of resulting in serious disease (Table 1) (41, 47). Huong *et al.* (2014) studied 159 *D. gallinae* samples from 142 farms in Japan using polymerase chain reaction (PCR) to indicate whether any of seven pathogens were present. They detected Avipox virus DNA in 22 samples (19 wild-type (ie. phenotype that occurs in nature)), 16S ribosomal RNA of *Mycoplasma synovia* in 15 samples (8 wild-type) and the *mgc2* gene of *Mycoplasma gallisepticum* was detected in two samples (identical to vaccine sequence) (48). It should be noted that whilst pathogen nucleic acids have been isolated from *D. gallinae*, this is not proof of

transmission.

	Pathogen	Association	Reference
Virus	Fowlpox virus	Transmission demonstrated	*(49)*
	St. Louis encephalitis	Isolated from mites	*(50-52)*
	Tick bourne encephalitis	Isolated from mites	(53)
	Eastern equine encephalitis	Transmission demonstrated	*(54)*
	Western equine encephalitis	Transmission demonstrated	(55)
	Venezualan equine encephalitis	Transmission demonstrated	*(56)*
	Avian leucosis	Unknown	*(57)*
	Newcastle disease	Isolated from mites	*(58)*
Bacteria	*Salmonella gallinarum*	Isolated from mites	(30, 59)
	Pasturella multocida	Transmission detected	(60)
	Erysipelthrix rhusiopathiae	Isolated from mites	(61)
	Coxiella burnetii	Transmission detected	*(62)*
	Listeria monocytogenes	Isolated from mites	(63)
	Mycoplasma synoviae	Isolated from mites	(48)
	Nocardia brasiliensis	Isolated from mites	(64)

Table 1: *Viral and bacterial pathogens 'associated' with Dermanyssus gallinae, Table based on information originally published by Valiente Moro et al., updated with information from Huong et al. (48) and edited from George et al. (65)*

1.4.7 Impact of *D. gallinae*

1.4.7.1 Economic impact

Dermanyssus gallinae causes significant economic loss to the European poultry industry, with estimates of ~230 million euros lost per annum (66, 67). This cost is largely attributed to higher feed conversion ratios, production losses and the cost of control (22). Estimated annual costs resulting from damage from *D. gallinae* for the UK alone were estimated at €3 million euros 11 years ago, whilst in Japan annual economic losses are estimated at €66.85 million euros (22, 68). Wójcik et al. (2000) estimated that excluding the costs of treatment, farmers lost ~36,000 Polish Zloty (~£6824) in a single production cycle and attributed this to increased mortality, decreased egg production and a reduction in the productive cycle (37). In 2001, Cosoroaba reported a statistically significant decrease in egg production in their study (69). Another study focusing on production losses in laying hens during infestation demonstrated a significant increase in the number of eggs laid, egg mass, feed intake, egg weight, percentage first choice eggs and body weight in heavily infested hens post treatment . They concluded that this confirmed that *D. gallinae* infestations result in an impact on the main performance traits which relate to profitability of laying hen farms, including the hens' general condition (70).

1.4.7.2 Production system impact

Research has shown that the production system utilised can impact *D. gallinae* infestation, with the egg laying industry predominantly affected due to a longer production cycle, resulting in a higher impact from *D. gallinae* when compared to the broiler industry, which utilises a short production cycle (19, 71). It has been noted that caged systems used by the layer industry could potentially provide an increase in potential hiding places (e.g. on egg conveyor belts or on transportation cages) resulting in avoidance from chemical control, thus exacerbating the problem (25). However, in egg laying production, research has demonstrated greater populations of *D. gallinae* in free-range systems compared to intensive, or caged, production systems. In the UK it has been demonstrated that between 60% (26) and 85% (24) of commercial egg-laying systems are affected by *D. gallinae* infestation. *Maurer* (1993) studied *D. gallinae* infestations in 39 poultry houses in Switzerland and noted that density of mites were higher in deep-litter systems which lacked a dung pit when compared to systems where dung storing facilities (a board or dung pit) and scratching areas were separated (36). Worldwide prevalence varies more, ranging from 20-90% depending on the production systems taken into consideration and the country. On a global scale, increased parasite population numbers seen in free-range system appear to be dependent on country, with factors such as farm and flock size being important governing factors influencing infestation rates (22).

1.4.7.3 Veterinary significance

As previously stated, *D. gallinae* has been shown to display plasticity in terms of host specificity, which means it is theoretically possible for *D. gallinae* to cause veterinary related issues to other species. It should be noted that whilst these report show presence of *D. gallinae* on the host, they do not always provide evidence of permanent infestations by *D. gallinae and* could be the result of opportunistic feeding. It has been reported to impact companion birds, including budgerigars, hobby pigeons and canaries, where in canaries it has been linked to *Chlamydia psittaci*. A study by Circella *et al.* (2011) observed an aviary of canaries suffering from *C. psittaci* and a heavy infestation of *D. gallinae* demonstrated an association between examined mites and infection, and led the authors to hypothesise a role of *D. gallinae* in spreading *C. psittaci* amongst canaries (56). In the literature, there are several reports that suggest *D. gallinae* can feed from dogs and cats (57-59), being attributed to causing dermatitis in a 16-year old domestic horse (60), recovered during skin sampling for mange mites from goats (61) and from mice inhabiting poultry houses (62).

1.4.7.4 Medical significance

In humans, contact with *D. gallinae* can result in pruritic dermatitis, thus posing an occupational hazard for poultry staff (72). There has been an increase in the frequency of reports of gamasoidosis associated with *D. gallinae*, resulting fromwild birds or backyard chickens, from residential, hospital, office space and occupational settings (including hobby keepers) in recent years. There were six reports between 1936 (73-77), one between 1962-1987 (78) and eighteen between 1988 and 2013 (72, 79-86), predominately from residential settings. Although the current body of literature available on this topic is small, it confirms that ingestion of human blood alone (73, 87) can allow for a persistent infestation and that there this issue is of global scale due to geographically wide-spread occurrence (65).

1.4.8 Host location

Ectoparasites that lack a permanent association to their host, such as *D. gallinae*, perpetually face the challenge of locating a host. Achieving this is dependent on multiple factors relevant to both host and parasite, as well as the environment that the parasite must navigate to locate a host (35). Research on haematophagous arthropod host-seeking behaviour demonstrates the entire process as dynamic and complex (88). Typically, ectoparasites rely upon a multitude of host-related stimuli, each stimulus having variable importance dependent on the context in which detection occurs (89). Certain stimuli are considered more important for close-range host detection; others are involved in long-range host location, whilst some are only deemed important in conjunction with other stimuli. Carbon dioxide is thought of as a typical host attractant for a variety of ectoparasites, regardless that the behavioural response will vary with other factors involved (35).

It has been shown that *D. gallinae* can be activated by temperature gradients as low as 0.005˚C. Research by Kilpinen and colleagues demonstrated that *D. gallinae* is highly sensitive to small temperature changes and that starved mites are more responsive to a heat cue than freshly fed ones (90, 91). Under dark conditions, carbon dioxide (CO_2) elicits no change in terms of turning angle or walking speed from *D. gallinae*, however in light conditions *D. gallinae* will “freeze” in response to an increase in CO_2 levels (92). This latter response is thought to reflect a defensive strategy by the mite to avoid being eaten by the hen (35). Increases in CO_2 indicate proximity of the host, thus increasing the likelihood of mites being detected and consumed by the host. As a result, mites “freeze” to avoid detection and wait for subsequent vibrations that are indicative of hen movement before becoming mobile again (35).

1.4.9 Pheromone cues in *D. gallinae*

Thigmokinesis and pheromone cues are used by mites to aggregate together (92, 93). Location of the mite's host is thought to be achieved through a combination of chemical signals, vibration responses, CO_2levels and temperature stimuli (35, 90, 91, 94). After the blood meal, *D. gallinae* form aggregations of mixed developmental stages. As is the case in the related *Dermanyssus prognephilus*, thigmokinesis and pheromones are also thought to play a role in this (93). Under ideal conditions (high temperature and hosts readily available), mites can feed every second or third night with each blood meal taking approximately 30-90 minutes (95). However, after a few days of starvation, the mites will also come out to feed in daylight (96). When starved, the mites do not discriminate between birds and humans, although they usually feed only on birds (94, 97). At this point in the host location process, it appears that the mites are not relying on species-specific stimuli but probably on more general stimuli, such as CO_2, vibration or heat, which are common to most vertebrates (35).

A study focusing on the plumage and skin of chickens (known to contain *D. gallinae* attracting substances, using chromatographic separation and *in vitro* feeding techniques demonstrated that surface skin lipid components act as host markers. They showed purified fowl diol esters of fatty acids, which are secreted from uropygial glands, and a natural extract made from bird surface lipids, were just as efficient as a feed stimulant for *D. gallinae* mites (94). Koenraadt and Dicke (2010) used a Y-tube olfactometer bioassay to understand mite choice when responding to cues that relate to either the host (chicken) or conspecific mites. Results showed a strong preference for volatiles produced by conspecifics from both starved mites (85%) and fed mites (84%) in comparison to a clean air stream. They also demonstrated a significant attraction to 'aged feathers' (i.e. those that remained in litter for 3-4 days) which was not seen for 'fresh feathers'. When using a 2.5% CO_2 air stream (designed to mimic the air exhaled by chickens), attraction was observed for fed mites but inhibited attraction for unfed mites, which favoured volatiles from aged feathers instead. From their data, they concluded that a combination of aggregation pheromones and kairomones are involved in the mediation of *D. gallinae* behaviour (92).

1.4.10 Why is poultry red mite an increasing issue?

Literature concerning *D. gallinae* dates back as far as the 1960s, but recognition has only increased in Europe in the last few decades. Speculation over drivers which underlie the rise of *D. gallinae* include a shift in production systems utilised by the laying sector of the poultry industry, an increasing human population (prompting an increase in meat and egg production), climate change and changes in regulation concerning chemical control. Recent changes to house

systems under Directive 1999/74/EC in 2012 placed a ban on using traditional cages for poultry. The main aim was to improve laying hen welfare; however, the ban promoted a shift towards more complex environmental conditions for chickens, which in turn provides a favourable environment for proliferation of *D. gallinae (98)*. Changes in perception regarding animal welfare have promoted a shift towards free-range productions systems across many European countries. In the UK, free-range systems have doubled in the last 11 years to now represent 50% of production (see figure below). Typically, larger mite populations are observed in free-range systems compared to cage units (24, 26), thus, an increase in free-range production systems may have directly contributed to an increase in *D. gallinae* levels. Coupled closely with these issues, are the complications involved in efficiently controlling *D. gallinae*.

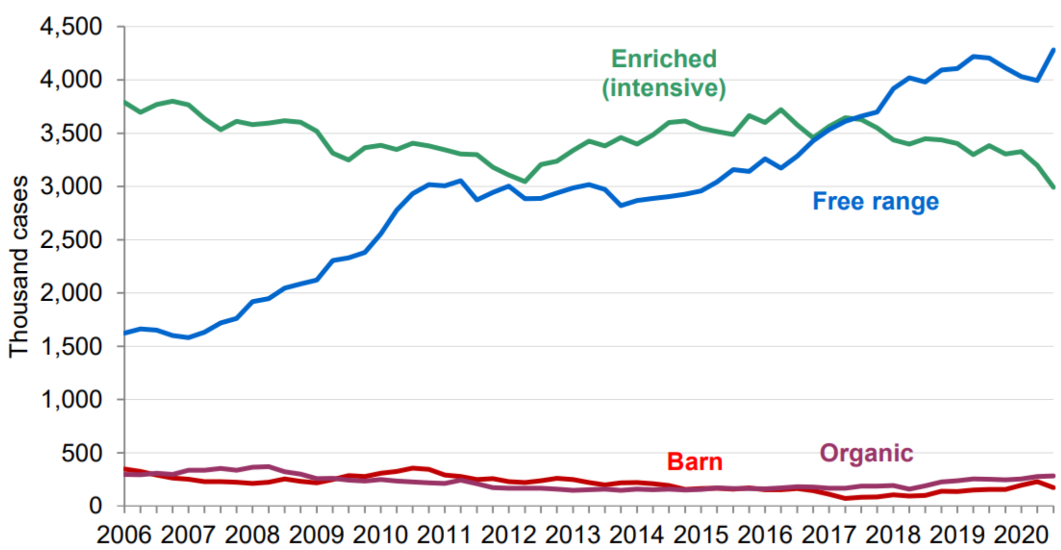


Figure 2: *Changes in egg production via differing production systems between 2006-2019 (99).*

1.4.11 Current control of *D. gallinae*

The fundamental basis for controlling *D. gallinae* is to follow strict rules of hygiene in the poultry system and to prevent introduction of the parasite from outside sources. After each production cycle, houses should be thoroughly disinfected and structure elements can be washed with warm water with the addition of additives lethal to mite eggs and oil-penetrating agents (20).

1.4.11.1 Monitoring D. gallinae populations

Monitoring levels of *D. gallinae* is essential to support effective control. Existing methods available for monitoring *D. gallinae* populations include perch traps (100), tubes containing fabric or cloth (36), corrugated cardboard or plastic traps (101), detection of *D. gallinae* in dust, feathers and impurities (102), examination of dried droppings (103), mite monitoring score (MMS) method (104) and an automated monitoring device developed by Mul et al. (2015) (105).

One major disadvantage to methods of monitoring mite levels is their labour-intensive nature, providing only a rough indication of population decline or growth in *D. gallinae* populations (106). As a result, farmers and egg producers do not frequently use them but typically apply 'traditional methods', such as passive and/or static traps (e.g. cardboard or plastic). However, without careful positioning of traps at multiple sites, which take into account *D gallinae* feeding behaviour and aggregation, underestimation of *D. gallinae* infestation can occur (105).

1.4.11.2 Conventional control measures

1.4.11.2.1 Acaricide use

Currently, control of *D. gallinae* primarily relies on use of acaricides such as organophosphates, carbamates, pyrethroids and formamidines (19, 107, 108). For example, Meyer-Kühling *et al.* (2007) showed a high efficacy for ByeMite ® (phoxim 50%, Bayer HealthCare, Animal Health division) against *D. gallinae* over a 49 day period, with efficacy exceeding 99% after day seven. In parallel a 400% increase in mite population size was witnessed in the control group (31). Another study by Keïta *et al.* (2006) found similar results, with 88.5% efficacy on day seven and >97% from day 10-49. However, increasingly strict regulations surrounding chemical use, lack of new acaricides and growing resistance to those that remain available, are limiting options for acaricide-mediated control of *D. gallinae* (33, 108). For example, in Sweden, no acaricides are currently licensed for use (109). Additionally, increased consumer awareness and greater demand for pesticide-free produce is helping to drive a move away from the use of chemical acaricides (110). In 2017, Exzolt was released by MSD Animal Health (14), using fluralaner as a novel control against *D. gallinae* through inclusion into the hens drinking water.

1.4.11.2.1.1 Presence of acaricide residues in produce

Concerns over the presence of pesticide residues in meat and eggs for human consumption are widespread and have been confirmed for acaricides in a study from Marangi *et al* (2012). Testing laying hens from Italy, Marangi and colleagues detected both permethrin and carbaryl acaricide residues. Specifically, a total of 225 samples were collected from 45 hens, and 91 (40.4%) of samples tested positive for carbaryl and four samples (1.7%) tested positive for permethrin. Carbaryl concentrations in the skin and fat of chickens were found to exceed the minimum detection limit (0.005 ppm) but remained under the maximum residue limit (MRL), before carbaryl was banned (25). In one farm, carbaryl contamination of organs and/or tissues was found to be significantly higher, where residues in both muscle and skin samples exceeded the MRL. Additional studies have found similar occurrence of acaricide residues for chemicals such as propoxur, which has been detected at more than six times the legal limit in the EU in eggs from laying hens (111). In 2017, the 'fipronil egg scandal' hit the headlines when ~200 farms in

the Netherlands were locked down after eggs were found to be contaminated with the pesticide, with levels exceeding the EU recommendation of 0.72mg/kg (112, 113). The Netherlands is one of Europe's largest egg producers and contaminated eggs were found as far away as Hong Kong (112).

1.4.11.2.2 Acaricide resistance

Resistance to acaricides has been described in the published literature, for example resistance to permethrin has been reported in the Czech Republic (114), France (108), Italy (115) and Sweden (116). In 2004, a survey focusing on British farms revealed that more than 60% reported experiencing acaricide resistant infestations (24). It is anticipated that this situation has now worsened with a rising percentage of acaricide resistant farms due to product misuse (25).

Resistance development can be influenced by a multitude of factors including biology, ecology, and genetics (117). Possible adaptations that allow individual mites to survive an otherwise lethal dose are numerous and typically classified based on physiological and/or biochemical properties. They are usually divided into two main categories (1) decreased response mechanisms or (2) decreased exposure mechanisms (118, 119). In most cases, sensitivity of the acaricide target is altered due to point mutations varied metabolism of the acaricide before it can reach the target site or changes in major detoxification enzymes (examples include; glutathione-S-transferases and P450 monooxygenases, reviewed in further detail in (120-123)).

1.4.11.2.3 Desiccant dusts

Physically acting substances such as oils and dusts represent an alternative to other control measures, such as acaricides (124). A number of desiccant or inert dusts are currently marketed across Europe for use in the control of *D. gallinae* populations and are frequently used as primary control agents (125). Desiccant dusts (DEs) and silica powders result in death of arthropods through desiccation, theorised to occur as a consequence of cuticle abrasion and absorption of cuticular lipids (126). They are capable of killing *D. gallinae* within a few hours or days (127, 128). As a group, desiccant dusts typically include a number of naturally occurring substances such as synthetic silica products, different minerals and DE products (126, 129). Products marketed for the control of *D. gallinae* are commonly modified by the addition of synthetic silica components, although some use purely synthetic products (128, 130). Desiccant dusts are reported to be most efficient at lower levels of relatively humidity (126, 128, 129, 131). When comparing different groups of DEs the typical rule is that pure DEs are less efficient than modified DEs, which in turn are less efficient than purely synthetic amorphous silicas. This is known to be correlated with the oil absorption capacity that is dependent on the particle size distribution of each type of

dust (129). Schulz *et al.* (2014) used twelve products, nine in powdered form and three in liquid form, to study efficacy of control against *D. gallinae*. They discovered that the mean time exposed to the product required for death varied significantly between products with 5.1-18.7 hours for powdered products and 5.5-12.7 hours for liquid products. The level of silicon dioxide content had no significant impact on efficacy; cation exchange capacity was positively related to efficacy, and water absorption capacity was negatively relatively to efficacy (132). One drawback of DEs is that they can result in the formation of dust during application that is detrimental and the decrease in efficacy as a result of high humidity can be problematic (130). Predicted costs of dust controls and spraying are €4.33/100 birds in caged systems and €3.83/100 birds in alternative systems (133).

1.4.11.3 Environmental control methods

One physical method to eliminate *D. gallinae* involves the application of high temperatures to empty poultry houses for several successive days. A gradual increase to a temperature of at least 45°C is achieved over the first day, followed by maintenance of this temperature for several days and finally a gradual decrease (20, 66). The main disadvantage to heat treatment is the high expense as well as the risk of heat-related damage to poultry house equipment (68). Using specific lighting programmes can aid in the control of *D. gallinae* populations as indicated by research from Belgium, where they found a light schedule of fifteen minutes light and forty-five minutes dark could promote a reduction in *D. gallinae* infestation (134). However, EU-Directive 1999/74 for the protection of laying hens states that a continuous dark period of at least eight hours must be provided and as such this light schedule would not be permitted in Europe (68).

1.4.11.3.1 Restricting movement and traps

One other option as an alternative to conventional acaricides is the development of an attract-and-kill strategy. This has shown promising results with other agricultural pests, such as codling moths, fruit flies and banana weevils (135-138). A major advantage is the reduced amount of the killing agent required and that it can be targeted to specific locations. This results in lower exposure and hazard to chickens and the environment. It will also slow down the development of resistance to chemical components. Ideally, biological kill components that are more environmentally friendly, such as essential plant oils and entomopathogenic fungi (138-140), can be incorporated in this strategy.

Chirico and Tauson (2002) utilised three different size traps containing 2% metriphonate in two separate trials, placed in areas mites are known to aggregate to assess efficacy. After a two-week trial a 99% reduction of *D. gallinae* was observed and after an eight-week trial a 95%

reduction of *D. gallinae* was observed in the placed traps. They noted that appropriate placement of treated traps was essential for adequate control (109). A similar study by Lundh *et al.* (2005) used cardboard traps that contained 20% neem oil (azadiracthin) placed in a layer farm that utilised a floor system, containing ~2,400 hens. Treated traps were used for four weeks, with traps replaced every week. At the end of the four weeks a 92% reduction in *D. gallinae* occurrence was observed. The successful use of traps to control *D. gallinae* populations relies heavily on the identification of mite aggregation sites, however, correct placing can avoid exposure of active ingredients to both birds and eggs and thus reduce the risk of acaricide residues in poultry products (141). Pritchard *et al.* (2016) conducted a study focused on restricting the movement of *D. gallinae* using a range of products and demonstrated containment of *D. gallinae* within a specified area (78-88%) when using barriers of double sided sticky tape, thyme oil and insecticidal glue (142).

1.4.11.4 Emerging and future control measures

1.4.11.4.1 Predatory mites

Use of predatory mites has been considered and attempted as a means of biological control. Maurer (1993) discovered the presence of the predatory mite *Cheyletus eruditus* (Schrank) in the litter of poultry houses. They demonstrated that the mites fed on juvenile *D. gallinae* and noted that when offered a choice between *D. gallinae* females and larvae, female *C. eruditus* exhibited a preference for larvae (36). Similar observations about *C. eruditus* have been reported in Mexico, Egypt and the UK, see: (143-146)). Lesna *et al.* (2009) investigated candidate predators through a survey of the mite fauna of European starlings (*Sturnus vulgaris*). They identified *Hypoaspis aculeifer* and *Androlaelaps casalis* as genuine predators of *D. gallinae* and noted other species that might act as predators including: *Hypoaspis miles*, *Cheyletus eruditus* and *Blattisocius keegani*. They caution that these species have not all been tested for the ability to feed on *D. gallinae* (147). A later study by Lesna *et al.* (2012) focused on *Androlaelaps casalis* and *Stratiolaelaps scimitus*, conducting small scale tests under three different temperatures; 26°C, 30°C (constant day and night) and 33-25°C (day-night cycle). Results showed that *A. casalis* produced better control of *D. gallinae* than *S. scimitus* for all trials, but complete eradication of *D. gallinae* was not achieved in any experimental condition. They hypothesised that this could be related to a prey refuge effect as the majority of predatory mites can be found residing in and around manure trays at the base of cages, whilst most *D. gallinae* individuals were located in higher areas of the cage (such as perches, nest boxes etc.) (148). Another study by Ali et al. (2012) compared four predatory mite species; *Hypoaspis miles* (Berlese), *Hypoaspis aculeifer* (Canestrini), *Phytoseiulus persimilis* (Athias-Henriot) and *Amblyseius degenerans* (Berlese). They found that *Hypoaspis* mites were the most effective predators of *D. gallinae* and completed

further experiments on *H. miles,* assessing the effect of physical, environmental and biological factors on efficacy. Results showed predation under all conditions, but included evidence of temperature-dependency and reduction in predation when alternative prey was provided (149).

1.4.11.5 Novel acaricides

1.4.11.5.1 Biopesticides

Spinosad is a natural product derived from the fermentation of *Saccharopolyspora spinosa* (150) and has been reported to have several characteristics which make it favourable for use as a pesticide, including natural degradation in light and the fact it doesn't bio-accumulate, persist or become volatile in the environment (151). George *et al.* (2010) researched the acaricidal activity and residual toxicity of spinosad and proved through *in vitro* testing that spinosad is toxic to adult *D. gallinae* females. The results demonstrated that both spinosad application and time post spraying had a significant effect on mite mortality ($P<0.001$), although, mite morality failed to reach 100% in any of the experiments.

1.4.11.5.2 Plant-derived products

Utilising the acaricidal effects of plant products has become increasingly popular (152). A range of mechanisms are thought to interfere with ectoparasites due to secondary metabolites in plant preparation, such as phenols, terpenes, flavonoids and/or cardiac glucosides (20, 153, 154). Previous work has focused on identifying natural plant preparations that are effective for use against *D. gallinae* including essential oils (155-158) and oriental medicinal plant extracts (159). Maurer *et al.* (2009) conducted a 168-hour *in vitro* experiment focusing on survival rates and reproduction of fed *D. gallinae* females and the effectiveness of four oils, seven plant preparations, one soap and three silicas. They demonstrated that all the oils and soap tested had a significant impact on *D. gallinae* survival, all silicas were effective at inhibiting reproduction but that only two of the plant preparations produced statistically significant control of *D. gallinae* through reduced survival and reproduction (124). Ghrabi-Gammar *et al.* (2009) focused on wild-growing plants from Tunisia, screening seven essential oils and commercial *Thymus vulgaris* for comparative measure. Their results showed yield of essential oil ranged from 0.1-0.5% dependant on the source plant and noted similar variation in toxicity results. Three essential oils studied did not cause a significant change in *D. gallinae* mortality when compared to the control. All other essential oils resulted in mortality rates that were statistically similar to mortality incurred using commercial *Thymus vulgaris* (90%), with *Pelagornium graveolens* oil resulting in a 100% mortality rate (160). Na *et al.* (2011) studied toxicity of cassia oils (two), cinnamon oils (four), (E)-cinnamic acid and (E)-cinnamaldehyde alongside structurally related compounds as potential acaricides against of *D. gallinae* using

vapour-phase mortality bioassay. Both cassia oils and all cinnamon oils demonstrated fumigant toxicity, as was also observed in six other compounds. Studies of structure-activity relationships revealed that structural characteristics (e.g. functional groups) rather than vapour pressure parameters could play a role in determining toxicity (161). Gorji *et al.* (2014) studied the field efficacy of garlic extract against *D. gallinae* in layer farms of Babol, Iran. They demonstrated that administration of garlic extract was efficacious at a 96.5% success rate against *D. gallinae* at selected mite-infested locations but noted that re-spraying was required to achieve even higher efficiency (162). Nechita *et al.* (2015) studied ten essential oils, focusing on their repellent and toxic effects. Lavender produced the highest mite mortality (over 97%), followed by thyme (84%) in filter-paper toxicity bio-assays. Post application effects (15 and 30 days) were also recorded, where thyme produced 100% mortality at 72 hours, followed by lavender (almost 80% after 72 hours). In terms of repellence, thyme (~80%), oregano (60%) and lavender (40%) exhibited the strongest effects, with all other oils showing efficacy lower than 30% (163). Tabari *et al.* (2015) considered the activity of carvacrol, thymol and farnesol. Carvacrol and thymol were demonstrated to be toxic against *D. gallinae,* producing higher efficacy than the synthetic pesticide permethrin, but farnesol caused no mortality (164). A later study by Barimani *et al.* (2016) focused on using traps containing carvacrol. *In vitro* tests revealed a 1% carvacrol formulation (with an ethyoxlayed castor oil used as an emulsifier) was significantly toxic to *D. gallinae*. They found that traps impregnated with carvacol had high efficacy, showing 92% reduction in mite populations after two weeks (165).

1.4.11.5.3 Use of entomopathogenic fungi

A recent consideration for control of *D. gallinae* is the use of entomopathogenic fungi (140, 166). A number of hypocrealean fungal species have been identified with known efficacy against arthropod pests and are currently used in agriculture and forestry (167). Laboratory experiments conducted using entomopathogenic fungi species have also shown *D. gallinae* susceptibility (140, 166). Experiments focusing on *Beauveria bassiana* (Bals) Vuill. and *Metarhizium anisopliae* (Metch) Sorokin have demonstrated limited success on a limited scale against *D. gallinae* in field trials (168). It has been shown that death of *D. gallinae* caused by entomopathogenic fungi can take several days (166).

Steenberg and Kilpinen (2014) conducted a study focusing on the synergistic interaction between the fungus *Beauveria bassiana* and three types of desiccant dusts, as well as combinations of the control agents against *D. gallinae*. They demonstrated significant synergistic interactions between *B. bassiana* and each of the desiccant dusts and significantly higher mite mortality rates than expected from an additive effect; as much as 38% higher. They also demonstrated this synergistic interaction when different methods of fungal application were

used at different levels of relative humidity. The speed of lethal action was not influenced by combining the two components (125). More broadly, in other pest species several studies in showed synergistic action between *B. bassiana* and different desiccant dusts, for example against beetles in stored produce (e.g. (169-171)) and mole crickets (172). Luz et al. (2012) demonstrated that a combination of diatomaceous earth and *M. anisopliae* resulted in a significant increase in efficacy against blood sucking triatomines (Triatoma infestans Klug). They also found that further combination with an oil produced an even higher mortality rate (173).

The mechanisms underpinning the synergistic interaction of entomopathogenic fungi and desiccant dusts are not well understood at present, however, several theories have been put forward. Desiccant dusts could result in lipid absorption from the host cuticle, thus making fungal infection more likely due to potential removal of fungicidal or fungistatic acids on the cuticle surface (169). The abrasive nature of desiccant dusts could allow nutrients to leak through the cuticle layer and promote the germination of the fungus conidia (174) as well as enhancing the adhesion ability of the conidia to the cuticle (171).

1.4.11.5.4 Attract and kill strategy

The development of an attract-and-kill strategy is one viable option as an alternative to conventional acaricide use, showing promise for other agricultural pests such as codling moths, fruit flies and banana weevils (135-138). A major advantage of this strategy is that smaller amounts of the killing agent are required, which can be applied to targeted locations. This results in lower exposure and risk to chickens, as well as slowing down the development of resistance to chemical components. Including essential plant oils and/or entomopathogenic fungi that are more environmentally friendly can strengthen the strategy (138-140). Due to *D. gallinae* having a concealed lifestyle, consisting of residing in cracks and crevices of the poultry house, during daytime it is difficult for pesticides to come into direct contact with the mite. Therefore an attract-and-kill strategy that effectively ‘brings the pest to the pesticide’ could have a higher chance of being effective (92). Mul and Koenraadt (2009) investigated preventing the spread and introduction of *D. gallinae* in poultry facilities by using the Hazard Analysis and Critical Control Points (HACCP) system. They stated that more in depth quantitative and epidemiological studies should be completed to aid effective control of mites on farms, but more importantly they noted that more studies are required that focus on mite behaviour including aggregation and host-seeking (68).

1.5 VACCINATION

Despite investment of millions of dollars and decades of research development of vaccines against parasitic infections remains relatively unsuccessful, especially for ectoparasites (175). Many factors have contributed to this, including the fact that parasitic infections are often chronic in nature. This chronic nature can be attributed to parasites frequently eliciting ineffective and/or inappropriate immune responses in their host or dampening the host immune system, resulting in prevention of an effective and/or robust immune response (176). Many parasites follow complex lifecycles that can complicate the process of developing efficacious vaccines. Additionally, many parasites demonstrate numerous immune evasion strategies such as molecular mimicry, antigenic variation and/or sequestration at both the infective and individual population levels (177). For ectoparasites, the development of vaccines can be even more difficult where direct interaction with the host is limited to invasive feeding.

Control of arthropod ectoparasites is an integral part of livestock management (182). Despite the enormous importance of ectoparasites in disease transmission (183) and economic loss (184), only one vaccine has been brought to market (TickGARD™), reflecting the difficulty in identifying suitable antigenic targets. Vaccination offers advantages including prolonged efficacy, freedom from chemical residues, reduced environmental pollution and reduced risk of resistance (186). It has been suggested that vaccines to blood-feeding ectoparasites may result in effective and sustainable control (187-189).

1.5.1.1 *Current status of arthropod vaccines*

A quarter of a century has passed since the first and only vaccine against arthropod ectoparasites was registered and commercialised for the control of cattle tick infestations (178). Research has since focused on the discovery of new protective antigens via various methodological approaches in a number of tick species (179-181) as well as other ectoparasites (e.g., (182, 183), including antigens such as Subolesin/Akirin protective against multiple ectoparasites (184) . In a 2019 editorial de la Fuente and Estrada-Pena, discussed why new vaccines for controlling ectoparasites have not been registered and commercialised, collecting data from PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and the European Patent Office (https://www.epo.org/searching-for-patents.html) to provide a comparison of the number of publications referring to antigens derived from ectoparasites with the total number of publications and patents which appeared in 2017-2019 (Figure 3) . They demonstrated that the majority of patents and publications relate to vector borne pathogen derived antigens and in relation to work on mite species, publications typically addressed mite induced allergies (185).

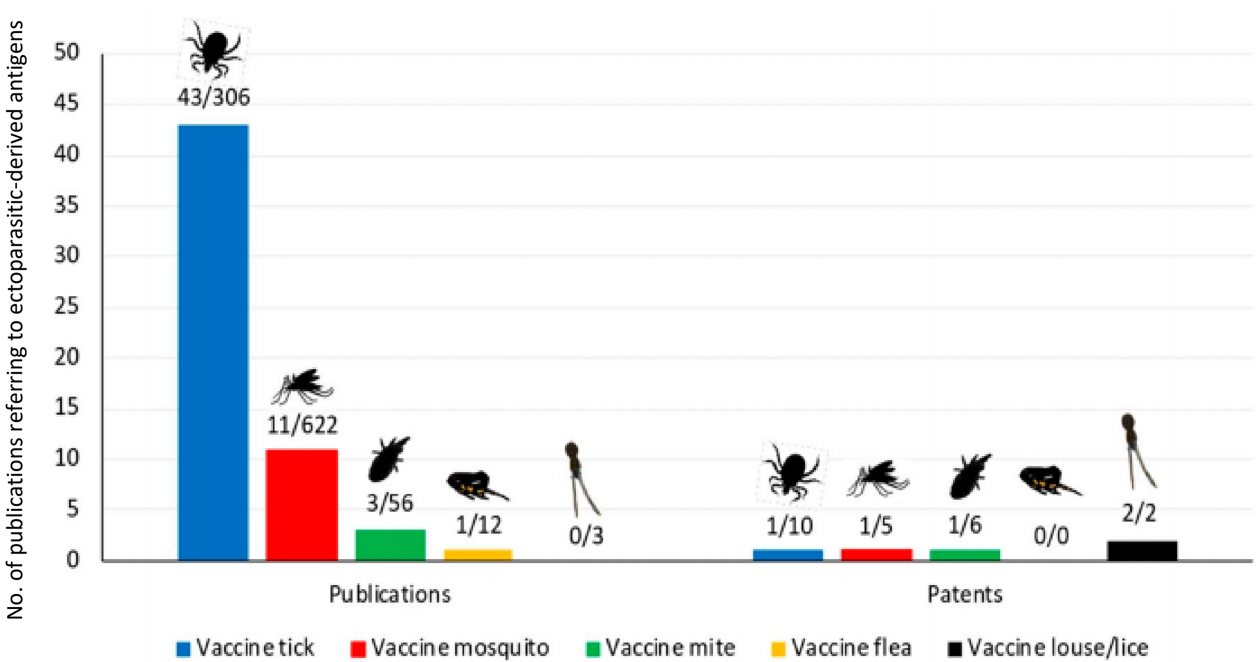


Figure 3*: Current status of arthropod ectoparasite vaccines, reproduced from (185). The graph depicts the number of publications and/or patents which refer to vaccines using ectoparasite-derived antigens over the total number of publications and/or patents which appeared between 2017- 2019*

1.5.2 Natural infestation of *D. gallinae* and hen immune response

All motile *D. gallinae* lifecycle stages (larva, protonymph, deutonymph and adult), except larva, are hematophagous (96), and as such, all are potentially exposed and susceptible to the effects of ingested antibodies present in the blood of vaccinated hens (186). Previous studies have demonstrated that vaccination of hens is a feasible strategy for controlling *D. gallinae* (186-192), however, studies focusing on mite population levels in poultry flocks suggest that birds are not capable of forming protective immunity to the mites following natural exposure (193). When hosts are infested with ectoparasites the host comes into contact with the parasites "exposed" antigens (194) but the immune response to "exposed" antigens of *D. gallinae* seems to ineffective at controlling populations of mites. Recent research has demonstrated that *D. gallinae* is capable of both modulating host inflammatory response (191, 195) and adapting reproductive behaviour to the host (195).

Avian IgY is thought to be, despite essential differences between molecules, the precursor of mammalian IgE and IgG. A study by Arkle *et al.* (2006) used enzyme-linked immunosorbent assays (ELISA) and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to assess the immunological response of humoral antibodies (derived from blood sera and egg yolk) to naturally occurring *D. gallinae* mite antigens. Results demonstrated no significant relationship between serum IgY level and IgY production or between serum IgY level and *D. gallinae* population level, however, they observed a numerical association when comparing mean values of mite population and antibody levels. They also found mean levels of total yolk-IgY to be significantly higher than serum-IgY, but stated that this could be the direct result of the extraction process used (38). Harrington *et al.* (2009) used ELISA and western blotting to study antibody response to immunization with immunoglobulins extracted from egg yolk. They observed a significant difference in IgY response in hens immunised with *D. gallinae* compared to controls, but no difference in IgM between groups. They observed a 50.6% increase in mite mortality after hens were immunised with soluble proteins extracted from *D. gallinae* through an *in vitro* mite feeding model (196). Market et al., (2016) immunized hens with various *D. gallinae* protein extractions which had been formulated using two different adjuvants and then isolated IgY-antibodies from hen eggs. They used an *in vitro* feeding assay, where they spiked chicken blood with Ig-Y preparations to detect antibodies that might contribute to *D. gallinae* mortality. In comparison to control mites, *in vitro* feeding of *D. gallinae* mites with IgY isolated from immunized hens (with one of the adjuvants) showed a significant increase in mortality (157).

1.5.3 Vaccination against *D. gallinae*

Currently, research into vaccine candidate identification for *D. gallinae* has used a 'rational' approach, which identifies antigens that are homologous to protective antigens from other ectoparasite species (187), or a 'pragmatic' approach, which involves fractionating native protein extracts of mites (197). The use of a 'rational' approach has been most common and relies on two main requirements in order to be successful; (1) knowledge of molecules essential to ectoparasite survival and (2) demonstration of accessibility of these molecules to the host immune system (198).

Bartley et al. (2009) discovered in *D. gallinae* an orthologue of tick histamine release factor (HRF) and hypothesised a potential regulatory role in mites. They produced recombinant protein Dg-HRF-1 and completed in vitro testing which showed a significant increase of mite mortality in comparison to controls of 7% after just one blood meal (187). Bartley et al. (2015) used anion exchange chromatography to fractionate an aqueous extract of *D. gallinae*, used the fractions to immunise hens, and then fed blood from these individuals to mites to study mortality rates. They demonstrated that immunisation with Deg-VIT-1 (vitellogenin), Deg-HGP-1 (hemelipoglycoprotein), Deg-SRP-1 (serpin), or Deg-PUF-1 (unknown function) caused a significant increase in mite mortality, ranging from 1.7—2.8 times higher than controls (186). Another study by Bartley et al. (2017) compared vaccine efficacy for a native preparation of soluble mite extract (SME) and a recombinant antigen cocktail vaccine including the immunogenic SME proteins: Derg-SRP-1, Deg-VIT-1 and Deg-PUF-1. They conducted the trial in a commercial-style cage facility with laying hens. They demonstrated a 78% reduction in *D. gallinae* counts for birds vaccinated with the SME but saw no statistically significant difference in mean *D. gallinae* numbers using the recombinant antigen cocktail vaccine (182). Arkle et al. (2008) demonstrated no significant difference in the mortality of *D. gallinae* when fed a diet of blood from individuals that had been immunised using proteins (*D. gallinae*) and the controls. Wright et al. (2009) generated four protein fractions and injected five groups with one fraction each in QuilA adjuvant and one control group to generate antibodies. They utilised an in vitro feeding assay to feed mites fresh chicken blood (with antibodies against protein fractions) and observed variability in the number of mites feeding. They found the phosphate-buffered saline (PBS) protein fraction to have the highest efficacy in terms of mite mortality with the cumulative percentage mortality after 24 hours being 16.4% compared to 10.9% for controls. They observed a rise in cumulative percentage mortality after 14 days from 16.4 to 34.8% in the PBS treatment group, compared to 10.9% to 27.3% in the control group (199).

Nisbet *et al.* (2006) studied tropomyosin, which was shown to be ubiquitously present throughout the gut and body of *D. gallinae*. They also demonstrated that ingestion of IgY occurs when mites feed and suggested that it could be good candidate for use in a vaccine as a "concealed" antigen (i.e. antigens that are normally hidden from the host) (200). Wright *et al.* (2016) immunised hens with recombinant versions of paramyosin (Der g 11) and tropomyosin (Der g 10). They then studied mite mortality after feeding from the blood of immunised chickens and found a significant increase of mortality in *D. gallinae*, suggesting that they could be useful vaccine candidates (192).

Bartley et al. (2012) completed an assessment of Cathepsin D and L-like proteins of *D. gallinae* for use a potential vaccine antigens. Results demonstrated significantly higher mortality rates after mites were fed anti-*Dg*-CatD-1 compared to controls. Survival analysis found anti-*Dg*-CatD-1 and anti-*Dg*-CatL-1 IgY had 4.42 and 2.13 times higher risk of mites dying compared to controls, suggesting that both *Dg*-CatD-1 and L-1 have potential as vaccine antigens (188). Price et al. (2019) evaluated vaccine delivery systems for induction of long-lived antibody responses to *D. gallinae* antigen in laying hens. After characterising the candidate vaccine antigen *D. gallinae* Cathepsin D-1 (Dg-CatD-1), they used it as a model for three different vaccine delivery methods: (i) purified recombinant Dg-CatD1 in adjuvant (Montanide^{TM} ISA 71 VG); ii) recombinant DNA vaccination to express Dg-CatD-1 and; iii) oral challenge utilising live transgenic coccidial parasites, *Eimeria tenella* expressing Dg-CatD-1. The results from two independent trials showed that only one strategy produced a long-lasting and strong serum anti-rDG-CatD1 IgY response: - purified rDg-Cat-D1 in adjuvant. They also noted a significant reduction in egg laying rates of *D. gallinae* mites after vaccination (201).

Lima-Barbero *et al.* (2019) identified the Akirin (AKR) gene from *D. gallinae* (Deg-akr), completed the production of recombinant Deg-AKR and evaluated its efficacy as a vaccine candidate in controlling *D. gallinae*. They found a 42% reduction in mite oviposition following feeding on vaccinated hens and a negative correlation between serum anti-Deg-AKR IgY levels and mite oviposition, providing support for Deg-AKR as a candidate protective antigen. A second study by Lima-Barbero et al. (2019) utilised proteomic data from fed and unfed adult and nymph mites to identify a novel PRM protein, calumenin (Deg-CALU) as a potential vaccine candidate. They conducted an on-hen trial, using *Rhipicephalus microplus* Subolesin (Rhm-SUB) as a positive control and found a reduction in the number of fed females which laid eggs by 35% (Deg-CALU) and 44% (Rhm-SUB), and a reduction in the number of eggs laid per fed mite by 37% (Deg-CALU) and 52% (Rhm-SUB). Their study provides support for their use as candidate protective antigens for *D. gallinae* control (190).

1.5.4 Summary of current vaccine research against *D. gallinae*

A summary of the antigens tested as vaccine candidates for the control *D. gallinae* infestation is detailed below in Table 2, produced either using recombinant proteins (i.e. artificially manufactured), native proteins (i.e. extracted in their natural form) or as a DNA vaccine (i.e. containing the DNA sequence encoding the antigen) .

Antigen	Type	Species	Adjuvant	Test	Effects (%)	Reference
Soluble protein mite extract	Native	*D. gallinae*	Incomplete Freund's	In vivo (199)	↑ 0.1 M	(199)
Soluble protein mite extract			QuilA	In vitro (197)	↑ 24 M	(186)
IEX Group 4					↑ 23.5 M	(186)
IEX Group 5					↑ 11.4 M	(186)
IEX Group 2					↓ 4.2 M	(186)
IEX Group 1					↑ 19.5 M	(186)
IEX Group 3					↑ 13 M	(186)
PBS soluble mite extract					↑ 10.1 M	(197)
Membrane associated					↑ 2.2 M	(197)
Urea soluble					↑ 0.2 M	(197)
Integral membrane					↓ 1.5 M	(197)
Mite extract			ISA 50 V	In vitro (196)	↑ 50.7 M*	(196)
Soluble protein mite extract			ISA 207 VG	Field	↓ 78 Pop*	(182)
Akirin	Recombinant	*Aedes albopictus*	ISA 50 V	In vitro (196)	↑ 35.1 M*	(202)
Bm86		*Rhipicephalus microplus*			↑ 23 M*	(202)
Histamine release factor		*D. gallinae*	QuilA	In vitro (203)	↑ 4.1 M*	(187)
Cathepsin D-1				In vitro (197)	↑ 6.9M	(188)
Cathepsin L-1					↑ 2.6M*	(188)
Unknown function protein 1					↑ 18.4 M*	(186)
Unknown function protein 2					↑ 0.6 M	(186)
Aspartyl proteinase					↑ 5.6M	(186)
Phosphoglycerate dehydrogenase	Recombinant	*D. gallinae*	QuilA	In vitro (197)	↑ 4.1 M	(186)
Serpin-1					↑ 12 M*	(186)
Hemelipoglycoprotein-1					↑ 18.9 M*	(186)
Vitellogenin-1					↑ 21.9 M*	(186)

Peptidase C1A-like cysteine proteinase					↑ 14.5 M	(186)
Serpin-2					↓ 8.2 M	(186)
Unknown function protein 3					↑ 3.5 M	(186)
Paramyosin					↑ 20.1M*	(192)
Tropomyosin					↑ 16.5 M*	(192)
Deg-SRP-1 + Deg-VIT-1 + Deg PUF-1			ISA 70 VG	Field	-	(182)
Calumenin			ISA 71 VG	On hen (147)	↓ 35 O*	(190)
Akirin					↓ 42 O*	(189)
Cathepsin D-1					↓ 50 *	(201)
Subolesin		*Rhipicephalus microplus*			↓ 44 O*	(190)
Cathepsin D-1	DNA	*D. gallinae*	Chicken IL-21		-	(201)
Cathepsin D-1			*Eimeria tenella*		-	(201)

Table 2: *Summary of antigens tested as vaccine candidates to control infestation by D. gallinae. * The effects are significant. Abbreviations: M = Mortality, O = Oviposition, up arrow = increase, down arrow = decrease, IEX= Ion-exchange chromatography. Sourced from(204).*

1.5.5 Considerations for vaccinating against *D. gallinae*

The variation in efficacy response to the vaccines tested against *D. gallinae* and responses following natural infestation suggests the host immune system may not be responding to antigens that impact the parasite (205). Thus, antigen selection could involve antigens that are not usually exposed to the hen in order to mount an immune response that provides protection against *D. gallinae*. Other factors could also result in avoidance of the immune response, e.g. pharmacologically active salivary molecules, the epidemiology of the mite, and the challenges of providing immunisation to birds during lay and likely requirement for re-vaccination to ensure sufficient circulating antibody titre (206). Furthermore, *D. gallinae* populations may take anywhere from three to six months to become established inside a poultry shed (101, 193) and during the first months of the laying cycle reduced numbers could stimulate a negligible immune response in the host resulting in a drop in IgY titres. As a result, birds may be left with an inability to cope immunologically once there is a rapid increase in population size (196). A reduced *D. gallinae* population will only present as an issue if the protective antigens in question are "exposed" antigens, "concealed" antigens will have no effect on the drop off of antigen specific IgY levels regardless of the mite population size. One consideration for vaccinating against *D. gallinae* is the short time spent feeding, approximately 30-90 minutes (20). This presents an issue when considering the antibody levels present in the chicken post vaccination, if *D. gallinae* fails to feed when antibody levels are high then efficacy may be negatively affected, and booster vaccinations may be required. If a single epitope is critical for successful immunisation, the same pressures that select for resistance against antiparasitic drugs will be brought to bear on the

vaccine target. Narrowly defined subunit vaccines may in particular be susceptible to resistance development; this question has not been addressed experimentally (207). Finally, the cost involved in development, manufacturing, registering and providing marketing for a vaccine is extremely expensive, costing millions to tens of millions of dollars in Western countries for a veterinary vaccine (208).

1.6 SUMMARY OF CONTROL METHODS

Whilst many methods are available or in development, realistic control *of D. gallinae* is not currently aimed at eradiation of the mite. Rather a reduction of PRM to levels that are economically profitable to farmers and that are acceptable towards the well-being of housed poultry.

Due to an ever-increasing resistance of *D. gallinae* to acaricides and changes in both legislation and production practice worldwide, it is predicted that *D. gallinae* will continue to pose a significant threat to global poultry production (33). Effective and new control measures for *D. gallinae* are now important not only for poultry production, but also in other sectors such as public health (33).

1.7 STUDYING GENETIC DIVERSITY

1.7.1 Molecular markers

A number of molecular markers are commonly used in the study of genetic diversity in arthropods, including mitochondrial DNA (mtDNA), internal transcribed spacer regions (ITS1 and ITS2), various genomic microsatellites and mitochondrial cytochrome c oxidase subunit 1. A short overview of these is provided below.

1.7.1.1 Nuclear ribosomal DNA

Most phylogenetic studies on mites have used the nuclear ribosomal gene cluster, especially the second internal transcribed spacer region (ITS2). Three genes make up this cluster; 18S rDNA, 5.8S rDNA and 28S rDNA which undergo transcription into RNA but are never translated into a protein. ITS1 and ITS2 divide the 18S and 5.8S, and 5.8S and 28S rDNA genes respectively and have been used for understanding closer relationships, for example for studying multiple species in one genus, given their faster rate of evolution (209).

1.7.1.2 Mitochondrial genes

Mitochondrial genes are widely used in molecular systematics due (in part) to their high copy number when compared to single-copy number genes (210). They also show strict maternal inheritance and lack of meiotic recombination (211) which simplifies analysis and representation of within-species variation data (212). Disadvantages to using mitochondrial genes and mtDNA are associated with the complex mutation process of mtDNA, which is known to be variable both in time and space, as well affected by its highly oxidative and metabolic environment (212). Generally, mitochondrial genes are classified in two broad categories (1) protein-coding genes and (2) mitochondrial ribosomal genes (12S and 16S rDNA) (210, 213). Cytochrome c oxidase subunit 1 (COI) in particular, has been successfully used to characterise multiple species (214, 215) and is often considered the "barcode for life" (216, 217).

1.7.2 Utilising bioinformatics

Analysis of variant data resulting from genome or exome sequencing is valuable for progress in biology, from basic research to translational genomics (218). With the development and deployment of high-throughput sequencing platforms, DNA sequencing can now be generated at unprecedented rates by individual laboratories, enabling the collection of large amounts of genetic data (219, 220). However, these masses of data pose substantial challenges for downstream studies, as highly specialised software tools and expertise are necessary to analyse and interpret them (220). A variety of bioinformatics tools have been developed to perform different analytical steps, including the alignment of the raw sequencing reads to reference genomes (221) (such as Burrow-Wheelers Aligner (222) and Bowtie (223)), the assembly of new genomes (224) (such as FERMI (225), Abyss (226) and SoapDenovo (218)), the application of data quality control (220) and the calling of single-nucleotide variants (SNVs) (227) (such as the Genome Analysis toolkit (GATK) (228)) or structural variants (SVs) (229) (such as CNVnator (230) and ERDS (231)).

1.8 GENETICS OF *D. GALLINAE*

Despite the importance of *D. gallinae* as a major parasitic pest plaguing the poultry industry, and the opportunities offered by genetic approaches, until recently there has been a distinct lack of genetic studies or nucleotide data in public databases. More recently, a draft genome sequence assembly has become available (232).

1.8.1 Genetic diversity of *D. gallinae*

Studies of genetic diversity in *D. gallinae* to date have focused on the COI gene (233-237), 16S rDNA (18, 236), and the ITS regions (45, 233, 235-238). Overall, the results have implied international and intranational migrations of mites. Studies by Roy *et al.* (2009; 2010) investigating species limits for the Dermanyssidae included a number of mite species and numerous isolates of *D. gallinae* from multiple regions in Europe (18, 236). They discovered within-species variation of <9% and, in conjunction with further analysis, suggested that *D. gallinae* represents a complex of hybridized lineages, possibly species, from a total number of 35 haplotypes (236). Roy *et al*. (2010) used an intronic nuclear region in the *tropomyosin* locus which revealed phylogenetic signal within *Dermanyssus,* and they believe constitutes a new interesting marker for phylogenetic explorations of Mesostigmata and other arthropods. Their study revealed that EF-1 alpha is not an appropriate marker within the *Dermanyssus* genus due to duplication events (236).

1.8.2 Internal transcribed spacer regions

Brännstrom *et al.* (2008) studied genetic differences between *D. gallinae* from wild birds and domestic chickens collected from four different geographical regions of south-central Sweden. They amplified a fragment of ITS1-5.S-ITS2 locus from nineteen individual *D. gallinae* isolates which yielded similar results showing all sequences from domestic chickens to be identical, all sequences from wild birds to be identical, but that the ITS1 sequence differed between the two mite groups. They observed ten fixed differences to be present between the two groups, however ITS2 and the 160-bp 5.8S rRNA gene sequence were both identical across all mites used (45). A later study focusing on *D. gallinae* isolates collected predominantly from layer farms in Norway and Sweden (and ten samples from wild birds utilised in the previous study) found 100% identical ITS sequences, except one sample showing a C-T substitution (235).

In Italy, one study focusing on ITS1, 5.8S ribosomal DNA and ITS2 showed no variation in the 5.8S and ITS2 regions and minimal intraspecific variation involving both substitutions and deletions in the ITS1 region (238). A later study utilising 360 sequences for ITS+ (ITS1, 5.8S and ITS2) detected no insertions, deletions or any unusual mutations and phylogenetic analysis resulted in predominantly unresolved clusters (233). Similarly, Roy et al. (2010) demonstrated that ITS1 and ITS2 lack information when focusing at an intraspecific level and that at an interspecific level provided very little characteristics within *Dermanyssus* when compared to other mite groups.

In Korea, nuclear ITS regions were classified into three sequence types with an additional ITS sequence defined as an intermediate type, suggesting among Korean mite populations occurrence of a hybridisation event (239).

1.8.3 Cytochrome c oxidase subunit I (COI)

Marangi *et al.* (2009) researched the phylogenetic relationship of *D. gallinae* populations using a 365-bp fragment of the COI gene. They found that 51.5% of sites were conserved from thirty-four DNA sequences of *D. gallinae*, whilst 48.5% of sites were phylogenetically informative. They aligned four *D. gallinae* COI sequences from the UK and found 66 nucleotide variations to be present, where three samples presented the same nucleotide substitutions and one sample differed. Results from mites obtained in France showed 8.5% and 14.2% nucleotide variations and from Italy sample variation was observed to range from 6.3-7.7%. These results suggested that variation within the UK is greater than variation within both France and Italy, although the sample sizes were small. They hypothesised that this demonstrated differential selection pressures put on *D. gallinae* within different countries, a view that was strengthened by the observations of nucleotide changes and variation in polymorphisms present between two farms in the same region in France (234). Marangi *et al*. (2015) used 360 amplicons of a fragment of the COI gene from twenty-four farms in Italy and found no insertions or deletions and noted no unusual mutations (i.e. frameshift causing deletions and/or deletions or false stop codons). Phylogenetic analysis of 24 COI sequences obtained from Northern and Southern Italy clustered into two main groups (A and B), with 11 farms in group A and 13 farms in group B.

Molecular investigations using COI sequences from Norway and Sweden revealed 32 haplotypes but, in most cases, a single haplotype was found on individual farms, suggesting that transmission routes are limited in these countries and that infestations are recycled. Haplotypes were not shared between Norway and Sweden, despite finding Swedish and Norwegian isolates in the two major haplogroups, providing evidence for minimal or no recent exchange of mites between these countries and no link between haplotype and geographical location was established (235).

Chu *et al*. (2015) partially sequenced the COI gene from 239 *D. gallinae* isolates and found 28 haplotypes present in Japan, suggesting that mites of several haplotypes found to be distributed throughout Japan were genetically related to those from European countries (237). A later study by Oh et al. (2019) demonstrated similar evidence in Korea. They found all samples had an identical COI sequence, which had also been reported in Europe and Japan, and phylogenetic analysis showed mites from Korea were genetically related to those present in other countries (239).

1.9 HAPLODIPLOIDY IN ACARI SPECIES

Haplodiploidy is a system whereby production of male offspring is monopolised by the mother, either through asexual production (arrhenotoky) or through sons eliminating their father's genome from their germline (paternal genome elimination, PGE) (240, 241). It is currently believed that approximately 15% of arthropods utilise a haplodiploid system (240, 242, 243). When considering genetic systems, mites demonstrate considerable diversity (244). Besides diploidy in both sexes, there are a number of forms of male haploidy which differ in whether males arise due to fertilised eggs (pseudo-arrhenotoky) or unfertilised eggs (arrhenotoky). In several mite families thelytoky (i.e. parthenogenetic production of females) occurs sporadically, whilst some taxa consist solely of thelytokus forms (244).

1.9.1 Sexual selection under haplodiploidy

Sexual selection arises due to competition within a sex, typically males, for access to mates (and thus gametes) (9) which can produce exaggerated traits through evolution. These traits evolve when females choose to mate with males that possess them (245). This can be due to the trait signalling genetic quality (the handicap principle) (246), or because inheritance to their sons will promote them to be attractive to other females (Fisherian runaway selection) (9). Under haplodiploidy maternal grandfathers are the closest male progenitor to sons because they do not inherit traits directly from their father, meaning that selection on male traits skips a generation (247). A simulation study (248) demonstrated it is particularly likely that rare alleles encoding male ornaments will be lost through genetic drift as a result of this delay in inheritance. It is possible this is also occurs for alleles that underlie traits that increase a males' reproductive success (e.g. combat ability), even if they are not a direct target of female selection. Deterministic models also demonstrated that haploid transmission genetics impact the genetic correlation between female preference and male traits, causing a promotion of sexual selection by the handicap principle but hindering the Fisherian runaway selection (249). In combination, these models suggest that haplodiploidy will affect the evolution of exaggerated male traits (245).

1.9.2 Evolutionary genetics of haplodiploidy

The evolutionary genetics of a species can be influenced due to haplodiploidy. Arrhenotokous and PGE systems result in maternal and haploid gene expression in males (245). As a result, in males, recessive alleles are exposed to selection and therefore (1) genetic load is reduced, as a result of a lower effective mutation rate and exposure of deleterious alleles (250) and (2) spread of rare recessive beneficial mutations occurs at an increased rate. Consequently, species utilising

arrhenotoky or PGE systems are hypothesised to be to adapt quickly to changing environments for non-sex specific traits. Inheritance does not occur directly between father and son meaning that male-limited trait evolution is complex. Selection occurring in females will impact evolutionary change more significantly, due to the fact that each gene is present more frequently in females compared to males (251).

Under these assumptions, it is possible to hypothesise that *D. gallinae* will be capable of adaptation against control measures, including vaccination, with females adapting more quickly than males but selection in males will be absolute. Target vaccine candidates should be selected to be experiencing strong purifying selection, as genes under purifying selection are typically well conserved due to possessing an essential function. Targeting these genes makes adaptation more difficult for *D. gallinae* against vaccination due to the critical function they perform. This is a typical feature of concealed antigens due to little or no immune-led selection for diversity.

1.9.3 Selective advantages of arrhenotoky vs pseudo-arrhenotoky

Pseudo-arrhenotoky is considered, as far as is currently known, to be a relatively rare phenomenon when compared to arrhenotoky (252). However, it is thought to have independently evolved in arthropods at least three times and, due to less investigation and being harder to detect than arrhenotoky, it is possible that it is more widespread than first thought (252).

According to Bull (1979, 1983) there are a number of selective advantages to male haploidy. One such advantage being the twofold representation of maternal genes in haploid gametes when compared to diploid sons coming from a biparental origin. This means that the probability of gene identity via descent between grandmother and her grandchild is double for uniparental sons than biparental. The twofold advantage of uniparental son production may override the potential lowering of fitness in these sons and could be an important consideration in understanding why male haploid systems evolved (253, 254). This advantage lends itself equally well to both arrhenotoky and pseudo-arrhenotoky (252).

Following Bull (1983), a second advantage of arrhenotoky is that it allows mothers to control the sex of her offspring by influencing the fertilisation of each egg. Hence, there is a flexible mechanism in arrhenotokous organisms that can change the sex ratio in an adaptive way whenever investment in one sex is more profitable than investment in the other (254). However, this ability to control the sex ratio is now known to occur in pseudo-arrhenotokous phytoseiid mites as well (252, 255). The third advantage of arrhenotoky is that virgin females colonising uninhabited sites can produce sons, whereas virgin female's incapable of parthenogenesis

cannot. If her sons mature before she ceases reproduction, they can mate with her so that she produces daughters and establishes a population. Arrhenotoky might evolve for this reason, although so would any other form of parthenogenesis. However, under diplodiploidy as well as pseudo-arrhenotoky eggs of either sex must be fertilised, so that virgin females cannot produce offspring unless they find a male produced by another inseminated female. Clearly, a low density of mates selects for arrhenotoky at the expense of pseudo-arrhenotoky. Thus, pseudo-arrhenotokous organisms are only expected when the chance of remaining unmated is virtually zero.

1.9.4 Haplodiploidy in *D. gallinae*

Currently, *D. gallinae* is assumed to use an arrhenotokous system, due to a lack of evidence for use of any other system in the Dermanyssidae (256). A study by Oliver (1965) determined that *D. gallinae* had a haploid number of three and a diploid number of six chromosomes. They estimated the two isobrachial chromosomes to be ~7.0 and 6.8µ and the heterobrachial one to be ~5.4µ long. They noted that female *D. gallinae* did not oviposit and that mated females laid haploid and diploid eggs in a 1:1 ratio and produced progeny of both sexes. Close agreement between the results of rearing experiments and of chromosome analyses of eggs indicated that sex determination in *D. gallinae* was of the haplodiploid type (257). Cruickshank and Thomas (1999) studied the evolution of haplodiploidy in Dermanyssine mites by analysing a 751bp 28S rDNA fragment from a group of mites that includes arrhenotokous, pseudoarrhenotokous and members which are ancestrally diplodiploid. After applying maximum-parsimony, maximum-likelihood and neighbour-joining methodologies they showed that all methods indicated arrhenotokous members of the clade arose from a pseudoarrhenotokous ancestor as opposed to directly from a diplodiploid one (256).

1.10 Concluding remarks

Dermanyssus gallinae infestation results in significant welfare and economic impact to the chicken egg industry annually. Current control measures are mostly ineffective and widespread resistance has been demonstrated against the available acaricides. Novel control strategies are desperately required to reduce the impact of *D. gallinae*, including the development of a suitable vaccine or novel acaricides. Remarkably little is known about the population structure and genetic diversity of *D. gallinae*, including how populations respond to selection by acaricides or future vaccines. A greater understanding of *D. gallinae* population structure and genetic diversity is required to safeguard the efficacy of future novel control methods (including vaccines and acaricides).

1.11 AIMS, OBJECTIVES AND HYPOTHESES

Overall Aim: To improve understanding of *D. gallinae* population structure, genetic and antigenic diversity with relevance to the development of, and likely response to, novel control measures.

1.11.1 Objective 1: Sample of *D. gallinae* across the United Kingdom and Europe

This will be completed by collection of *D. gallinae* from layer farms to from the UK and Europe to cover spatial analysis, and sampling the same farms in the UK over time to cover temporal analysis

1.11.2 Objective 2: Identify and validate genetic markers for *D. gallinae*

This will be investigated through collection and processing of *D. gallinae*, genetic characteristion of individual *D. gallinae* from Europe and targeting a fragment of the cytochrome c oxidase subunit 1 (COI) gene, to gain insight into population structure and genetic diversity.

1.11.3 Objective 3: Conduct genome-wide genetic analysis to assess *D. gallinae* population structure and regional variation

This will be investigated through utilisation of bioinformatic tools and pipelines to use the *D. gallinae* transcriptomic resources for SNP identification against the draft *D. gallinae* genome assembly. This will be followed by (iii) assessment of the population structure of *D. gallinae* using a Mid-Plex genotyping assay for genome-wide genetic analysis of UK and other European *D. gallinae* populations.

1.11.4 Objective 4: Studying the occurrence of genotypes associated with acaricide resistance to pyrethroids

This will be investigated through assessment of the occurrence of VGSC genotypes that have been previously related to resistance to pyrethroid acaricides

1.11.5 Objective 5: Assessment of antigenic diversity at loci encoding anti-*D. gallinae* vaccine candidates

This will be investigated through gene-specific genetic analysis focusing on assessment of existing genetic diversity and signatures of selection at loci encoding prominent anti-*D. gallinae* vaccine candidates

2 General Methodology

2.1 Mite collection

2.1.1 Mite collection during visits to UK farms

Dermanyssus gallinae were collected directly from the environment during visits to four commercial egg-producing farms, two based in West Sussex, one in Oxfordshire, and one in East Sussex. All farms were known to have a history of *D. gallinae* infestation. Mites, and associated debris, were collected using a small metal spatula/spoon (15cm long x 0.22cm handle diameter; Sigma-Aldrich, St Louis, Missouri, US) to scrape them from cracks and crevices in and around hen nesting boxes or from the legs of the automatic feed conveyer belt into a 135mm x 135mm x 20mm weigh boat (Sigma-Aldrich, St Louis, Missouri, US). Typically, each nest box and legs were systematically scraped for *D. gallinae* following a logical order through the poultry house. After the weigh boat was ~30% full, mites were transferred into a Nunc^{TM} EasYFlask^{TM} 75cm^3 (ThermoFisher Scientific, Waltham, Massachusetts, US) with a vented cap firmly screwed on. Collection was repeated until a flask became about one quarter full, containing roughly 1-2g of mites. After the first mite collection was transferred to a flask, flasks were tapped on the floor before each subsequent opening to ensure that the mites already captured fell to the bottom of the flask and did not escape. Each collection lasted 4-6 hours during which between 10-20g of mites and debris was collected. Once in the laboratory, flasks were stored at room temperature upright on three wire test tube racks (26cm x 9cm x 13cm) inside a tray of soapy water (36cm x 15.5cm x 36cm) (5% fairy liquid (Procter & Gamble, Cincinnati, Ohio, US) in tap water) to prevent any mites escaping (Figure 4). After use all waste, including gloves, single-use overalls and any other disposable items (e.g. weigh boats) used as part of the collection was sent for autoclaving and disposal.

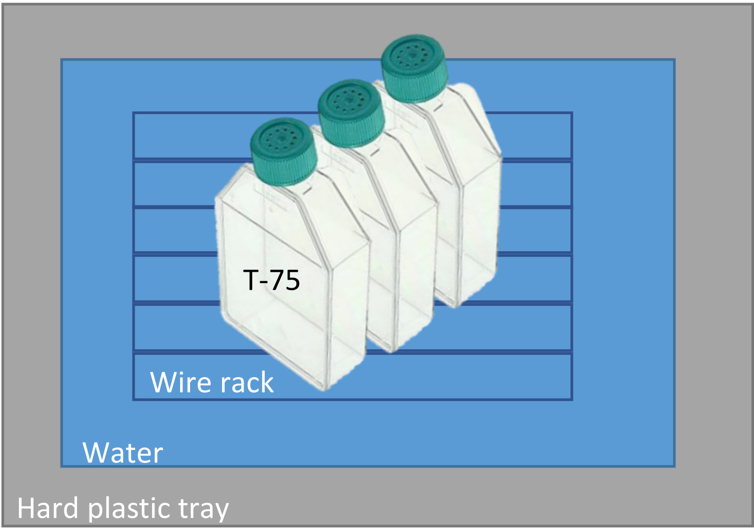


Figure 4: *An illustration of the set up used for storing poultry red mite after collection in T-75 flasks in the laboratory*

To separate mites from environmental contaminants and debris such as dust, faeces and mud, flasks were left at room temperature in T-75 flasks where they migrate towards the vented cap, aggregating inside the screw cap allowing for clean harvesting. Once a sufficient number were aggregated (i.e. the inside of the cap was completely covered by *D. gallinae* and no longer visible), the cap was carefully removed and screwed onto a fresh T75 flask. Each flask was tapped on a flat surface (e.g. lab bench) to dislodge mites to the bottom if the flasks then caps from both flasks were tightly re-sealed. This was repeated over a few hours until a layer of approximately 1cm of mites was visible when tapped onto the bottom of the T-75 flask. It was observed that storing too many clean mites in a single T-75 flask resulted in a faster mite death rate so additional mites were stored in separate flasks to improve survival. Mites were transferred to fresh flasks approximately every three days; if delayed, mite survival was reduced. To achieve this, flasks were tapped onto the laboratory bench to dislodge mites from the caps then the contents of the old flask tipped into the new one. Mites were either used immediately for experiments, frozen at -20˚C in a fed state or left to starve for seven days at room temperature with two flask changes (allowing for digestion of a blood meal and any chicken-derived proteins, outlined in (197, 203)) before being frozen at -20˚C.

2.1.2 Remote mite collection from UK farms

To increase number of mite collections and improve representation across the UK, additional farmers of laying hens were recruited for remote mite collection. Farmers were identified through personal recommendation by expert contacts, through Google and Facebook and contact with egg supplier companies. A convenience sampling frame was used and farmers were contacted directly by email, through facebook, or by telephone. All farmers received instructions on how to carry out collection (summarised in Figure 5, Supplementary 10.1.2), an envelope containing two labelled ziplock bags, two rolled traps and four flat traps (Figure 5), a questionnaire (Supplementary 10.1.1), an information sheet relating to the project (Supplementary 10.1.3).

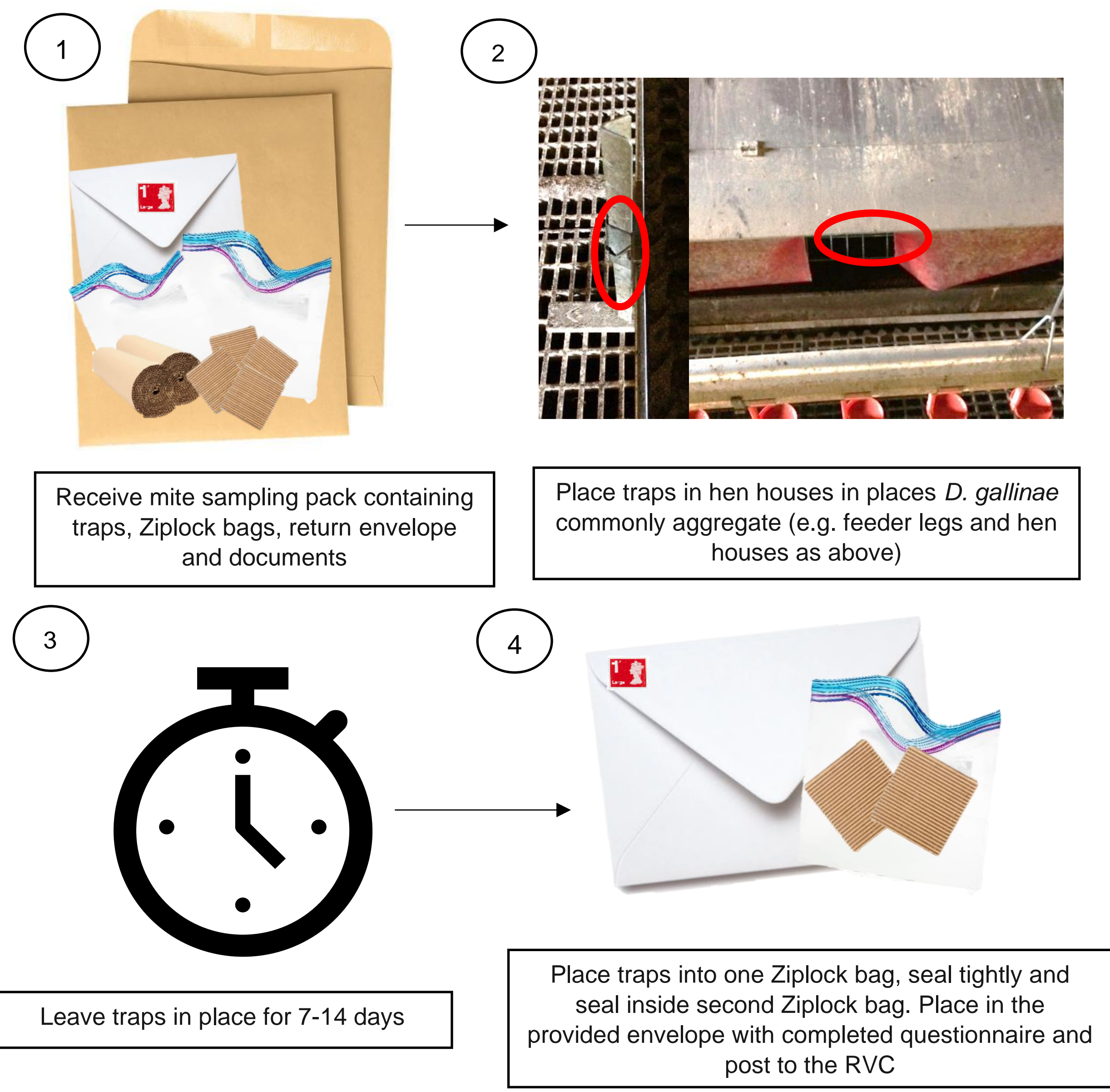


Figure 5: *Summary of sampling advice provided to farmers in four steps: 1) Receiving the mite sampling pack 2) Placing the traps around the poultry houses 3) leaving the traps for 7-14 days and 4) packaging traps into Ziplock bags to send to the RVC*

Briefly, farmers were asked to capture mites using cardboard traps as previously described in (101) . Farmers were asked to place the traps in positions inside their poultry houses common for *D. gallinae* infestation (such as Figure 5) for 7-14 days and then remove and seal them into two 26.8cm x 27.cm ziploc bags (S.C. Johnson & Son, Inc, Racine, Wisconsin, US) before posting to the RVC. Prepaid postage stamps were provided.

Upon arrival at the RVC, traps were removed and put into T75 flasks. The plastic ziplock bags were cut into horizontal strips, approximately 2cm wide, and placed into a second T75 flask as mites tended to aggregate in the corners and folds of the bags. Mites were either used immediately for experiments, frozen at -20˚C in a fed state or left to starve for 7 days (allowing for digestion of a blood meal and any chicken-derived proteins) before being frozen at -20˚C.

2.1.3 Remote mite collection from European farms

Mite samples from layer farms in mainland Europe were received from academic partners identified through the COREMI red mite European COST action (found at: https://www.coremi.eu/home.html). Samples had either been collected for previous research or were collected from farms as described above in 2.1.2 and were received at RVC preserved in 70-100% (v/v) ethanol in 1.5ml tubes or alive in cardboard traps. Mites were either used directly, dried and frozen at -20˚C, or preserved in ethanol (>70% v/v).

2.2 HANDLING *D. GALLINAE* IN THE LABORATORY

When transfer of live *D. gallinae* was required in a laboratory setting benches were sprayed with Total Mite Kill (Nettex, Rumenco, Burton-Upon-Trent, UK) containing 0.23% (w/w) permethrin. Gloves, bottles, spatulas and other equipment used were also sprayed with Total Mite Kill and washed in hot soapy water after use. Any waste was placed in two autoclave bags and autoclaved after transfer was completed.

2.3 ROUTINE DNA EXTRACTION FROM WHOLE MITES

DNA was extracted from individual and pooled mite (~50-100 mites) samples using a Qiagen Blood and Tissue Kit (Qiagen, GmbH, Hilden, Germany) with the following modifications. An individual mite or pools of mites were homogenised in 180µL of Qiagen kit ATL buffer where they were broken by slicing the whole body with a sterile Agani™ 21G x 1 ½'' (0.8 x 38mm) needle (Terumo, Tokyko, Japan)). 20µL of 20mg/mL proteinase K solution was added and the resulting suspension was mixed by vortexing for 60 seconds before samples were incubated at 56˚C overnight using a thermomixer – mixer HC (STAR LAB GmbH, Hamburg, Germany) at 450rpm. The Qiagen protocol was resumed from step two and DNA samples were eluted in 40-200µL of DNase and RNase free water (Invitrogen, Paisley, UK), dependant on final volume and concentration required for downstream analysis.

2.4 RNASE TREATMENT

After DNA extraction, 0.1µL of 10 mg/mL RNAse A (Thermo Scientific™), Waltham, Massachusetts, US) was mixed per 10µL of genomic DNA and incubated at 37 ˚C for one hour in the thermomixer – mixer HC (Star lab, Hamburg, Germany) at 350rpm.

2.5 DNA QUANTIFICATION SYSTEMS

For DNA quantification, a NanoDrop system was used for a preliminary assessment of DNA concentration and to detect the presence of contaminants (e.g. salt, protein). A Qubit system was used for accurate quantification of double-stranded DNA (dsDNA) concentration when required.

2.5.1 Nanodrop

DNA concentration and purity were determined using a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) following the standard manufacturers protocol. For aqueous solutions of nucleic acid, the default DNA-50 setting was used to measure 1µL of sample.

2.5.2 Qubit

Accurate DNA concentration was determined using a Qubit High Sensitivity (dsDNA) kit or a Qubit Broad range (dsDNA) kit on a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts), following the manufacturer's protocol with 1µL of sample DNA mixed with 199µL of Qubit® working solution.

2.6 PRIMER DESIGN AND USE

Primer (oligonucleotide) pairs were designed based on sequence data for each DNA target using the OligoEvaluator™ by Sigma-Aldrich (found at: http://www.oligoevaluator.com) (258) for use in Sanger Sequencing from genomic or coding DNA. Primers were designed to balance specificity with amplification efficiency, with design aimed to minimise secondary structure and the occurrence of primer dimers and achieve a melting temperature between ~55-72˚C and primer length of ~16-22 bases. A minimum primer length was set at 16 nucleotides due to primers < 16 nucleotides typically having lower annealing specificity to target DNA (259). Secondary structures were minimised as they are known to impede primer annealing and extension (259). All primers used for PCR reactions were synthesised by Sigma-Aldrich (St Louis, Missouri, US) and were delivered lyophilised. Stock concentrations of 100 um were made by adding DNase and RNase free water (Invitrogen, Paisley, UK) as per the volumes recommended by the manufacturer. From the stock primer, 1 in 10 dilutions were routinely prepared as working stocks, also using DNase and RNase free water (Invitrogen, Paisley, UK). Primers were stored at -20˚C prior to use. Primers for SNP genotyping by next generation multiplex sequencing were all designed by Eurofins Genomics.

2.7 POLYMERASE CHAIN REACTION (PCR)

PCR was routinely perfomed in a volume of 25µl containing 12.5µl 2× MyTaq™ DNA polymerase (Bioline, London, UK), 400 pmol of each primer (1 µl forward, 1 µl reverse) (Sigma-Aldrich, Darmstadt, Germany) and 2µl of DNA template (0.5ng-25ng), made up to volume with DNase and RNase free water (Invitrogen, Paisley, UK). PCR cycling conditions were initial denaturing at 95˚C for 5 minutes, followed by between 30-35 cycles of denaturing at 95˚C for 30 seconds, annealing at 55-65˚C for 30 seconds, and elongation at 72˚C for 30-60 seconds. Final elongation was performed at 72˚C for 5 min. A T-Gradient thermocycler (Biometra, Jena, Germany) or G-Storm Thermal cycler (G-Storm, Somerton, Somerset UK) were used to perform PCR.

2.8 TRIS-BORATE-ETHYLENEDIAMINETETRAACETIC ACID (TBE) BUFFER PREPARATION

2.8.1 Stock solution of TBE

A stock of 5X TBE buffer was produced by weighing 54g of Tris base (Formula weight (FW) – 121.14)(Sigma-Aldrich, St Louis, Missouri, US) and 27.5g of boric acid (FW=61.83) (Fisher Scientific, Hampton, New Hampshire, US) and dissolving both in 900ml of ultra-purified deionised water using a magnetic stirrer. After both were fully dissolved, 20ml of 0.5 M ethylenediaminetetraacetic acid (EDTA) (Fisher Scientific, Hampton, New Hampshire, US) with pH 8.0 was added and the total volume adjusted with distilled water to make a final volume of 1L.

2.8.2 Working solution of TBE

For agarose gel electrophoresis, a working solution of 0.5 × TBE was used. This was prepared by diluting the stock TBE (see 2.8.1) to a 1:10 dilution, i.e. 100ml of stock TBE was mixed with 900ml of ultra-purified deionised water.

2.9 AGAROSE GEL ELECTROPHORESIS

After amplifcation, PCR products were resolved by electrophoresis in 1.0% (w/v) agarose gels. 50ml of 0.5× TBE buffer was mixed with 0.5g of ultra-pure agarose powder (Invitrogen – Fisher Scientific, Hampton, US) and heated in a microwave oven for 60 seconds (or until completely clear) in a 250ml glass conical flask (Sigma-Aldrich, St Louis, Missouri, US). The mixture was cooled by running the conical flask under cold water for approximately 30 seconds before 0.01% (v/v) Safeview Nucleic acid stain (NBS Biologicals, Cambridgeshire, UK) was added and mixed thoroughly. Gels were poured, a 16 or 20 comb was placed, and then left to set for

approximately 20 minutes before being fully submerged in 0.5×TBE buffer in an agarose gel electrophoresis tank (Bioline, London, UK). 1µL of each PCR product was mixed with 5µL of 5× DNA loading dye (Bioline, London, UK) and pipetted into a well in the agarose gel. As a size control 5µl of GeneRuler 1kb plus ladder (Thermo Fisher Scientific, Waltham, Massachusetts, US) was used for comparison to assess product size. Electrophoresis was performed at 70V on a Hoefer™ PS300-B (Harvard BioScience, Holliston, Massachusetts, US) power pack for approximately 30 minutes and then visualised using a U:Genius gel documentation system (Syngene, Cambridge, UK).

2.10 PCR PURIFICATION

PCR amplicons of the anticipated target size were purified using a QIAquick PCR column purification kit (Qiagen, GmbH, Hilden, Germany) following the manufacturer's instructions. The basis of the protocol stems from DNA binding to a silica membrane when a high concentration of chaotropic salts are present (260). The column was washed as recommended by the manufacturer and purified DNA was eluted using 30-50µL of DNase and RNase free water (Invitrogen, Paisley, UK), dependent on final concentration required for downstream sequencing (variable from starting DNA concentration).

2.11 SANGER SEQUENCING

Direct Sanger (chain termination) sequencing of purified PCR amplicons was carried out by Eurofins Genomics employing the same primers as used in the initial amplification. 20µL of each eluted PCR product were sent with 20µL of primer (400 pM) allocated for every eight samples. Once received, sequences were assembled and curated using CLC Genomics Workbench 8.1.3 (Qiagen, Aarhus, Denmark). Curated sequences were assembled and/ or aligned using CLC workbench 8.1.3, dependant on requirements, with default parameters and final assemblies/alignments were manually curated to detect errors.

3 SAMPLING *DERMANYSSUS GALLINAE* IN THE UNITED KINGDOM AND EUROPE

3.1 INTRODUCTION

Infestations of *D. gallinae* have been documented worldwide with up to 90% of layer chicken premises reported to be infested, although published studies have been scarce (21-24, 261). Investigations into prevalence of *D. gallinae* infestations on layer farms have been conducted in countries including; Algeria (262), France (19), Italy (25), Japan (263), Poland (21), South Korea (261), Tunisia (28) and the UK (24, 26). Recognition of the impact caused by *D. gallinae* has significantly increased in the past few decades, with mite populations reaching epidemic levels in some parts of the world (98, 264). With further increases likely in the absence of effective control, many of these previously published estimates of prevalence may now be outdated.

3.1.1 Prevalence of *D. gallinae* in the United Kingdom

Prevalence estimates of *D. gallinae* infesting commercial egg-laying systems in the United Kingdom were last published in 2004-2005, with results indicating 60-85% prevalence (24, 26). Higher occurrence was typically reported in free-range systems compared to intensive cage and barn systems (24, 26). One study found that just 7% of responding layer chicken farmers claimed they had never had a *D. gallinae* infestation on their farm (26). Guy et al., (2004) found that 87.5% of the 29 farms that participated in their study had *D. gallinae* present, and that populations of both nymphs and adult *D. gallinae* were significantly higher ($P<0.05$) in free range in comparison to both barn and caged production systems (24).

3.1.2 Prevalence of *D. gallinae* in Europe

Unlike in the UK, no consistent prevalence trends have been reported across poultry systems in other countries in Europe with different countries showing varying prevalence across production systems (22). In 2009, Sparagano et al., reviewed prevalence and key figures for *D. gallinae* in poultry farm systems (22). Their study demonstrated the lowest prevalence in Norway at 23% for caged systems. In Denmark, prevalence was highest for free-range farms (68%) and lowest in caged systems (32%), with barn systems in the middle (50%). Interestingly, results demonstrated a lower prevalence in organic farms (36%) than free-range systems. Meanwhile, data from France showed a higher prevalence in organic systems (80%) compared to free-range (56%), barn (50%) and caged systems (72%). In their study, prevalence for Italy was estimated at 74.1% whilst, a study into Swedish flocks in 1995 revealed higher prevalence in alternative systems (21%) compared to caged systems (4%) and backyard flocks (19%) (23). Investigations into the severity, control, cleaning, biosecurity and prevalence of *D. gallinae* in Dutch farms revealed that more than 80% of farmers reported *D. gallinae* present on their farms. Their study also showed that farms utilising battery cages (i.e. without enrichment) experienced more problems with *D. gallinae*, applied their first control measure at an earlier time and repeated it

more frequently than farmers using other production styles. Investigations into the prevalence of *D. gallinae* in 24 farms in Portugal between August 2016-November 2017 revealed 95.8% of farms were infested. The study covered a total of 21 civil parishes in 15 municipalities in the region of Centro and Norte (29). An assessment of *D. gallinae* in three extensive farms from three locations in Cluj, Romania, showed a prevalence range of 57.5-72.5% across farms (265).

3.1.3 Aims and Hypotheses

The aim of this chapter was to establish the current prevalence of *D. gallinae* in layer farms in the UK, collect and process *D. gallinae* for subsequent genetic analyses, determine the current level of knowledge UK farmers have in relation to the control and infestation of *D. gallinae*, and evaluate farmers opinions on the current efficacy of control methods utilised.

3.2 Methodology

3.2.1 Mite collection

Information on mite collection for the UK and Europe can be found in General Methodology sections 2.1.1 (Mite collection during visits to UK farms) and 2.1.2 (Remote collection from UK farms) for farms from the UK, and 2.1.3 for European farms. Four UK farms were sampled in person (UK6, UK7, UK11 and UK12) based on contact information provided by RVC colleagues. The remaining 21 farms were sampled through colleagues working on *D. gallinae* in the UK, contact through Google, Facebook and/or egg supplier companies (Table 3).

Method of approach	Number of farms
Email via knowledge from RVC colleagues	4
Directly received from colleagues working on *D. gallinae* in the UK	6
Provided by BEMB (funding body) farmers	2
Egg supplier	4
Google search	11
Facebook	2

Table 3: *Method of approach for sampling of* D. gallinae *and the number of farms collected through each approach*

3.2.2 Ethical approval

Ethical approval was given by the Social Science Research Ethical Review Board at the Royal Veterinary College (URN SR2017-2018) for work pertaining to the questionnaire. Confidential information was kept secure and all results were presented anonymously.

3.2.3 Questionnaire

Questionnaires were provided to all UK farmers as part of the sampling pack they received during mite collection (see Supplementary 10.1.1). Information collected is summarised here

and will be used for interpretation of the genetic analyses outlined later in the thesis. A detailed epidemiological study of risk factors has not been undertaken.

3.3 Results

3.3.1 Sample distribution across the United Kingdom

From the UK, a total of 25 farms were sampled, primarily based on convenience (availability), covering all four countries and including 18 counties (Figure 6, Table 4). *Dermanyssus gallinae* were detected from traps or in person from 23/25 farms, indicating a prevalence of 92%. A total of 64% of farms sampled were from England, 16% from Scotland, 16% from Northern Ireland and 4% from Wales. One farm sent *D. gallinae* mites in traps but provided no other information so is referred to as anonymous.

Figure 6: *Approximate location of farms sampled for D. gallinae across the United Kingdom with 15 farms in England, four in Scotland, four in Northern Ireland and one in Wales*

Country	County	Production type	Organic Status	Farm no.	Mites	Questionnare data
England	Durham	Free-range	Non-organic	UK 1	Y	Y
	East Sussex	Free-range	Organic	UK 6	Y	Y
	East Sussex	Free-Range	Organic	UK25	N	Y
	Gloucestershire	Free-range	Non-organic	UK 2	Y	N
	Hampshire	Free-range*	Non-organic	UK24	Y	N/A
	Kent	Free-range	Non-organic	UK 5	Y	N
	Lincolnshire	Free-range	Non-organic	UK 14	Y	Y
	Oxfordshire	Free-range	Non-organic	UK 7	Y	Y
	Suffolk	Free-range	Non-organic	UK 11	Y	Y
	West Sussex	Free-range	Non-organic	UK 12	Y	Y
	Cheshire	Intensive	Non-organic	UK 4	Y	Y
	Shropshire	Intensive	Non-organic	UK 8	Y	Y
	Tyne and Wear	Intensive	Non-organic	UK 15	Y	N
	Tyne and Wear	N/A	N/A	UK16	Y	N
	Tyne and Wear	N/A	N/A	UK17	Y	N
	N/A	N/A	N/A	UK18	Y	N
Northern Ireland	Country Antrim	Free-range	Non-organic	UK21	Y	Y
	Tyrone	Free-range	Non-organic	UK 10	Y	Y
	Tyrone	Free-range	Non-organic	UK22	Y	Y
	Tyrone	Free-range	Organic	UK23	Y	Y
Scotland	Angus	Free-range	Non-organic	UK24	N	Y
	Peebleshire	Free-Range	Non-organic	UK20	Y	N
	Highlands	Free-range	Organic	UK 9	Y	Y
	Peebleshire	Intensive	Non-organic	UK 13	Y	Y
Wales	Cardiganshire	Free-range	Organic	UK 3	Y	Y

Table 4: *Location of farms sampled in the UK including the farm number, county it resides in, the style of production system used, whether* D. gallinae *isolates were successfully received and if questionnaire data was complete. * indicates a free-range backyard flock. N/A = not available.*

3.3.2 Temporal collection from UK farms

To understand whether genetic diversity of *D. gallinae* changes over time, temporal sampling was completed at three free-range UK farms (UK 6, 7 and 11; Table 5).

For farms UK7 and UK11 only one barn was present and was consistently sampled across all three points. For UK6, three barns were present on site, barn one was consistently sampled, barn two was sampled twice and barn three was sampled once (Table 5), providing some information on intra-farm variation.

Country	County	Farm no.	Production type	Dates sampled	Season
England	Suffolk	UK 11	Free-range	12-02-2018	Winter
				07-12-2018	Summer
				29-05-2019	Spring
	Oxfordshire	UK 7	Free-range	30-09-2017	Autumn
				10-07-2017	Summer
				30-07-2017	Summer
	East Sussex	UK 6	Free-range	Barn1 29-05-2017	Spring
				Barn 1 07-06-2017	Summer
				Barn1 30-07-2017	Summer
				Barn 1 31-09-2017	Autumn
				Barn 1 13-11-17	Autumn
				Barn 2 13-11-17	Autumn
				Barn 3 13-11-17	Autumn
				Barn 1 09-08-2018	Summer
				Barn 1 31-08-2018	Summer
				Barn 1 15-01-2020	Winter
				Barn 2 15-01-2020	Winter

***Table 5**: Temporal sampling of three farms (UK11, UK 7 and UK6), county farm resides in, production type and dates of sampling provided*

3.3.3 Sample distribution across Europe

Across Europe, samples were received from 16 countries encompassing a total of 84 farms (Figure 7, Table 6). The average number of farms per country was five, with the highest number of farms sampled in Italy (17) and the lowest in Germany (1) and Turkey (1).

.

***Figure 7**: Map showing the origin of all* D. gallinae *populations sampled from 16 European countries*

Country	Closest Town or Region	Sample name(s)	Sampling date or year	Production system (if known)
Albania	Lushnye	ALB1	2018	(Broiler) Intensive
	Berat	ALB2	2017	(Broiler) Rural farm
	Korca	ALB3	2018	(Broiler) Rural farm
	Peshkopi	ALB4	2018	(Broiler) Intensive
	Durres	ALB5	2017	(Broiler) Rural farm
	Shkodfer	ALB6	2018	Broiler
Belgium	Destelbergen	BEL1	2018	Intensive (Layer)
	Destelbergen	BEL2	2018	Intensive (Layer)
	Destelbergen	BEL3	2018	Intensive (Layer)
	Evergm	BEL4	2018	Intensive (Layer)
	Evergm	BEL5	2018	Intensive (Layer)
	Evergm	BEL6	2018	Intensive(Layer)
Croatia	Zagreb	CRO1	28-09-17	Growing layer hens
	Zagreb	CRO2	28-09-17	Intensive (Layer)
	Zagreb	CRO3	22-03-18	Intensive (Layer)
	Zagreb	CRO4	22-03-18	Intensive (Layer)
	Zagreb	CRO5	2017	Intensive (Layer)
Czech Republic	Bohemia	CZH1	2018	Intensive (Layer)
	South Moravia	CZH2	2018	Intensive (Layer)
Denmark	Vejle	DEN1	-	-
	Jylland	DEN2	-	-
France	Lacepede	FRA1	-	Aviary (Layer)
	Montfaucon	FRA2	-	Intensive (Layer)
	Grenade sur Garonne	FRA3	-	Intensive (Layer)
	Saint-Pons-de-Thomieres	FRA4	-	Intensive (Layer)
Germany	Hannover	GER1	2018	University flock (Laying hens)
Greece	Thessaloniki	GRC1	2018	Intensive (Layer)
	Corinth	GRC2	2018	Intensive (Layer)
	Leros	GRC3	2018	Intensive (Layer)
	Attica	GRC4	2018	Intensive (Layer)
Italy	Ravenna	ITA1	16-01-18	Laying hens
	Padua	ITA2	23-01-18	Laying hens
	Lecce	ITA3	15-03-18	Laying hens
	Lodi	ITA4	23-02-18	Laying hens
	Cremona	ITA5	23-03-18	Laying hens
	Bologna	ITA6	03-03-18	Laying hens
	Bologna	ITA7	06-04-18	Laying hens
	Ravenna	ITA8	02-05-18	Pullets
	Lecce	ITA9	23-07-18	Laying hens
	Lecce	ITA10	20-07-18	Laying hens

	Varese	ITA11	27-07-18	Laying hens
	Bari	ITA12	08-08-18	Laying hens
	Verona	ITA13	10-05-18	Laying hens
	Brindisi	ITA14	21-04-18	Laying hens
	Verona	ITA15	18-03-18	Laying hens
	Milano	ITA16	29-03-18	Laying hens
	Pavia	ITA17	09-05-18	Breeder
Macedonia	Skojpe	MAC1	2018	Intensive (Four farms combined)
Netherlands	Flevolanol	NET1	-	Intensive (Layer)
	Flevolanol	NET2	-	Intensive (Layer)
	Flevolanol	NET3	-	Intensive (Layer)
	Flevolanol	NET4	-	Intensive (Layer)
	Flevolanol	NET5	-	Intensive (Layer)
	Lutten	NET6	2018	Intensive (Layer)
	Barneveld	NET7	2018	Intensive (Layer)
	Barneveld	NET8	2018	Intensive (Layer)
	Aalten	NET9	2018	Intensive (Layer)
	Unknown	NET10	2018	Intensive (Layer)
Portugal	Riveria	POR1	2018	Intensive (Layer)
	Benaveute	POR2	2018	Intensive (Layer)
	Braemes	POR3	2018	Intensive (Layer)
	Souta da Carpalhosa	POR4	2018	Intensive (Layer)
	Zezeroro	POR5	2018	Intensive (Layer)
	Oliveina de Fnades	POR6	2018	Intensive (Layer)
	Ancogelo, Panti de linne	POR7	2018	Intensive (Layer)
	Rego, Debasto	POR8	2018	Intensive (Layer)
Romania	Tatarlaua	ROM1	2018	Backyard
	Tatarlaua	ROM2	2018	Backyard
	Cuzdrioara	ROM3	2018	Backyard
	Cuzdrioara	ROM4	2018	Backyard
	Cuzdrioara	ROM5	2018	Backyard
	Floresti	ROM6	2018	Backyard
Slovenia	Tenetiše	SLO1	2019	Backyard
	Škofljica	SLO2	2019	Backyard
	Kamnik	SLO3	2019	Laying farm
Spain	Seville	SPA1	2018	Intensive (Layer)
	Seville	SPA2	2018	Intensive (Layer)
	Seville	SPA3	2018	Intensive (Layer)
	Seville	SPA4	2018	Intensive (Layer)
Turkey	Karacaali	TUR1	2018	Intensive (Layer)

Table 6: *Sample locations from Europe (outside of the UK), including the country and number of individual mites sampled*

3.3.4 Epidemiological data collection via questionnaire

From the UK, 18 farmers provided completed questionnaires. Summarised results are presented in Table 7 for husbandry related questions and Table 8 for *D. gallinae* specific questions, below.

Topic	Variable	Category	Number of farms	% of all total farms sampled
Production system	Production system	Free-range	16	88%
		Intensive	2	12%
	Organic status (Free-range farms only)	Organic	6	37.5%
		Non-Organic	10	62.5%
Housing	Housing system	Enriched cage	2	11%
		Multi-tier	3	17%
		Single tier	3	17%
		Battery	0	0%
		Flat deck	7	38%
		Free-range	3	17%
	Manure system	Manure belt	5	28%
		Manure pit	10	56%
		Deep litter	4	22%
		Other	1	6%
	Date of establishment (one unknown)	1960-1980	3	18%
		1981-2000	8	47%
		2001-2020	6	35%
	Number of poultry units/farm	1	5	28%
		2	4	22%
		3	4	22%
		4	1	6%

		5	1	6%
		6 or more	3	17%
	Access to wooden or metal structures for free-range hens	Wooden	6	37.5%
		Metal	4	25%
		Both	4	25%
		None	2	12.5%
Bedding	Type of litter	Shavings	9	50%
		Wood pellets	1	6%
		Woodchip/Easy chick	4	22%
		Straw	4	22%
		Nature Dug	1	6%
		N/A	2	11%
	Change in litter	Yes	5	28%
		No	11	61%
		N/A	2	11%
Chicken	Total number of chickens	1000-50,000	14	78%
		>50,000-100,000	2	11%
		>100,000	2	11%
	Age of chickens (one farm unknown)	<1 year	12	71%
		>1 year	5	29%
	Chicken breed	Hyline	12	67%
		Cotswold Legbar	1	6%
		Novagen White Leghorn	1	6%
		Novagen Brown	1	6%
		Lohman	2	11%
		Burford brown	1	6%
		British blacktail	2	11%
	Chicken supplier	Reared on site	2	11%
		CFP	2	11%
		Tom Banor	1	6%

		Country Fresh	2	11%
		CPH	2	11%
		Shea Eggs	4	22%
		Hinchliffes	1	6%
		JSR Services	1	6%
		Hayward	1	6%
		Hillside	1	6%
		Humpreys pullets	2	11%
		Stonevate	1	6%
Feeding	Type of feed utilised	LAF	1	6%
		ForFarmers	2	11%
		Home mix	2	11%
		Crown feed	1	6%
		272m free range layer	1	6%
		Standard layer	1	6%
		Harbird layers meal	1	6%
		Meal/Mash	6	33%
		Humphreys	2	11%
		Organic layer I	1	6%
	Use of feed additives for hen health	None	10	56%
		Yes	1	6%
		Biomoss	3	17%
		Sodium bicarbonate	1	6%
		Vitamins	1	6%
		Marigold seeds	1	6%
		Actigen	2	11%
		Arbosol	1	6%
		Marine shell	1	6%

***Table 7**: Summary of husbandry-related questionnaire data from 18 UK farms sampled for* D. gallinae *in relation to production system, housing, bedding, chickens and feeding, with the number of farms and % of total farms sampled provided. Questions permitted the respondent to answer with multiple categories causing discrepancies with total % not equalling 100%.*

Topic	Variable	Category	Number of farms	% of all total farms sampled
Dermanyssus gallinae	Are you in contact with local farmers? If so, are you aware if they are affected by *D. gallinae*?	Yes, affected	4	22%
		Yes, not affected	2	11%
		No contact	10	56%
		No answer	2	11%
	Have you noticed *D. gallinae* on your farm?	Yes	18	100%
		No	0	0%
	When did you first notice *D. gallinae* on your farm?	Less than a week	0	0%
		One week – one month previously	1	6%
		One month – six months	3	17%
		Six months or more	10	56%
		Can't remember	1	6%
		Always	3	17%
	Is *D. gallinae* infestation a significant problem on your farm?	Yes, significant	12	66%
		Yes, moderate	4	22%
		No	1	6%
		No answer	1	6%
	Do you believe there is sufficient information available about *D. gallinae*?	Yes	9	50%
		No	9	50%
	Do you notice a correlation between the age of chicken and level of *D. gallinae*?	Yes	13	72%
		No	5	28%
	Do you notice seasonal fluctuations in *D. gallinae*?	Yes	12	67%
		No	6	33%
	Signs associated with *D. gallinae* infestation	Visible mites	10	56%
		Perches and nest boxes	*2*	*11%*
		Crevices	*1*	*6%*
		On the eggs	*2*	*11%*

		Drinker lines	*1*	*6%*
		Central areas	*5*	*28%*
		Traps	*1*	*6%*
		Staff infested	1	6%
		Pale eggs	1	6%
		None	4	22%
	What major problems do you consider to be associated with *D. gallinae* infestation?	Production drop	15	83%
		Irritation	2	11%
		Blood loss/Anaemia	3	17%
		Staff infested	1	6%
		Feather loss	3	17%
		Disease and/or bacterial/viral transmission	5	28%
		Pale/Wrinkly eggs	4	22%
		Increased cleaning and control measures	2	11%
		Egg cleanliness	3	17%
		Mortality	5	28%
		Bird welfare	1	6%
		Pale combs	1	6%
Current control of *D. gallinae*	Current control measures	Chemical	8	44%
		Desiccant dusts	12	67%
		Hygiene	6	33%
		Temperature	0	0%
		Natural	5	28%
		Feed additives	0	0%
		Other	4	22%
	Number of control measures in use	One	7	39%
		Two	6	33%
		Three	4	22%

		Four	1	6%
	Method for cleaning empty poultry houses	Dry clean	3	17%
		Power wash and disinfectant	14	78%
		Fumigate	2	11%
		N/A	3	17%
	Do you feel your current control measures are effective?	Yes	11	61%
		No	7	39%
	Considerations for using a new control measure	Cost	7	39%
		Ease of application	6	33%
		Other	2	11%

Table 8: *Summary of questionnaire data defining opinions from 18 UK farms sampled for* D. gallinae *in relation to* D. gallinae specific *topics including current control measures, with the number of farms and % of total farms sampled provided. Questions permitted the respondent to answer with multiple categories causing discrepancies with total % not equalling 100%. Italics indicating number of farmers reporting specific locations of visible mites*

3.3.4.1 Production systems utilised

In total, 88% of farms sampled were free-range compared to 11% intensive. None of the farms sampled utilised a barn system. Of the free-range farms, almost two thirds were non-organic, with just 37.5% having organic status (Table 7).

3.3.4.2 Type of housing system and manure system used

Out of the 18 farms, the largest proportion used a flat deck housing system (39%) and a manure pit for waste management (50%) (Table 7). Two farms selected more than one option of manure style, choosing both manure pit and deep litter, explaining the difference in totals between graphs. Three farms answered with 'free-range' when asked which housing type they utilised rather than selecting from the available options. For manure style, one farm answered 'other' stating that they used 'two thirds slatted, one third scratch'.

3.3.4.3 Proximity to other farms and contact with local farmers

Farmers reported close proximity to other local farms housing poultry (<4 miles away) with just one farm reporting a distance of 10 miles. Seven farms reported a distance of less than 0.5 miles to the closest farm with two of these just 0.25 miles distance. The average distance to the next poultry farm was 2.3 miles. When asked if they had contact with local farmers and knowledge regarding whether these farms also had *D. gallinae* infestations, 56% of farmers reported that they had no contact, four farms reported that local farms were also affected and two farms reported that local farms had no *D. gallinae* at the time of sampling (Table 8). Two farmers did not respond to the question.

3.3.4.4 Questions relating to **D. gallinae** *infestation*

All farmers responded that they noticed the presence of *D. gallinae* on their farm. Awareness of when *D. gallinae* was first present on the farm ranged from the previous week to always, with one farmer reporting that they could not remember. Overall, all but one farmer felt that *D. gallinae* is a problem, with 67% responding that it is a significant issue and 28% responding it is a moderate issue. There was an almost even 50/50 split amongst farmers regarding whether they believe there is sufficient information available (Table 8).

Almost three quarters of farmers (72%) answered that they noticed a correlation between the age of their flock and the level of *D. gallinae* present on their farm, with one farmer noting that 'the number of mites decreases with the age of the chicken'. Meanwhile 67% of farmers noticed a seasonal fluctuation in *D. gallinae* infestation on their farm.

3.3.4.4.1 Signs associated with a heavy infestation of *D. gallinae*

The most commonly associated sign to a heavy infestation of *D. gallinae* was mite visibility, with 56% of farmers reporting (Table 8). Broken down into smaller categories, centralised areas within poultry barns were most common for visibility. One farmer noted that they inspect hen houses two hours after the lights have been turned off and that is when mite visibility is highest. One farmer utilised traps for visualisation of *D. gallinae* to monitor infestation levels. Two farmers also remarked on mite aggregation in relation to being visible, with one farmer stating that they 'group together in places where there is warmth and food'. Four farmers simply reported that there were no signs related to a heavy infestation, and one farmer reported that heavy infestations impacted their staff, stating that 'staff get covered in them'.

3.3.4.5 Major problems associated with D. gallinae infestation

A drop in egg production was the most commonly identified problem associated with *D. gallinae*, reported by 83% of farmers (Table 8). One farmer noted that poor egg production was 'due to red mite compounding other stress factors'. Egg cleanliness or pale/wrinkly eggs were noted by seven farmers, all of which contribute to production losses. Three farmers recorded factors relating to egg cleanliness, with one farmer noting that 'mites were present on eggs which required extra cleaning'. Bird welfare was only indicated directly by one farmer; however, irritation, blood loss/anaemia, feather loss, pale combs and mortality were all indicated as major problems and can be considered as welfare issues. Hen mortality was the highest welfare related issue, being reported by 28% of farmers, with one farmer noting that 'reduction in hen health' is a major problem and 'therefore mortality of hens caused by *D. gallinae*'. Hen mortality can also be linked to production losses. Two farmers commented that *D. gallinae* impacting on their staff were a major problem. One farmer stated that 'staff get covered and take them home', and another farmer noted that presence of *D. gallinae* is 'not nice for staff or hens'.

3.3.4.6 Current control measures utilised on farm

Desiccant dusts were the commonly used method of control across farms with 67% of farmers reporting their use (Table 8). Chemical control measures were utilised by almost half of farmers (44%), but it should be noted that chemical control is not permitted for organic farms. Four farmers selected "other" as a control measure, with two farmers using predatory mites, one farmer using a disinfectant (stating Interkokast) at hen turnover time, and one farmer stating use of detergents and an on/off lighting regime. No farmers reported use of temperature or feed additives as a control measure. Most farmers reported use of one control measure (39%), 33%

used two control measures, 22% reported a combination of three measures and one farmer reported using four control measures at the time of sampling.

3.3.4.7 Cleaning of poultry houses between flocks

Power washing with disinfectant was the most noted method of cleaning empty poultry houses between flocks (78%). From the farmers specifying disinfectant, most did not provide a brand name, but two farmers specified using Biogel or Viricia and one farmer specified using Vircon S. One farmer noted that they hired contractors for power washing. Three farmers dry clean their barns between flocks, with one stating they never wet wash and usually blow dry. Two farmers stated the use of fumigation. One farmer commented that they 'sometimes fumigate' and the other farmer noted that they starting fumigating at the end of the batch of chickens using 'Furaldehyde and permost', the latter a microencapsulated permethrin product.

Three farmers provided a time frame for when they clean their poultry house rather than a cleaning method, so were classified as not applicable. Fourteen farmers provided just one cleaning method and four farmers used two or more.

3.3.4.8 Effectiveness of current control measures

Over half of farmers (61%) felt that their current control measures were effective at controlling *D. gallinae* infestations on their farm. 39% of farmers felt that the current control measures were not effective (Table 8).

One farmer noted that 'control measures suppress the problem but don't eradicate the issue'. Another farmer commented that '100% control is nearly impossible to manage. Once under control I now just keep spraying, it seems as though if I can keep it so I just see adult mites and not any nymphs then I have it under control'. Another farmer that answered yes stated that his control methods only remained effective if he persevered in using them.

3.3.4.9 Considerations for using a new control measure

A total of 39% of famers reported that cost is a consideration for investing in a new control measure on their farm and just over a third (33%) considered ease of application as an important consideration for a new control measure. Nine farmers selected the 'Other' category, with three stating that they are limited by organic status. One farmer reported a 'lack of information', one felt that 'current control measures work best' and another reported 'Exzolt', a new fluralaner-based product (266). Three farmers commented on efficacy of products, with one remarking that they 'want to use something that is sure of working'. One farmer stated that they had contemplated the use of diatomaceous earth products but felt that the control was insufficient

to justify the cost, even though they did not select the cost option in their answers. Lastly, one farmer commented that 'red mite has to be continually treated for' and related to the fact 'resistance builds up quickly' and if you use 'expensive products you can spend a lot of money'. They also noted that spraying the front of the hen cages is easy, but the mites move to the back of the cages where it is harder to reach with control measures. Three farmers did not select any option for the question.

3.4 DISCUSSION

Overall, *D. gallinae* mites were collected from a total of 25 UK farms covering all four countries and including 18 counties. Questionnaire data was received from 18 UK farms, with 16 having *D. gallinae* mites successfully received. *Dermanyssus gallinae* were also sampled from 16 other European countries, representing 84 farms, with 1-17 farms per country included.

3.4.1 Prevalence of *D. gallinae* across the UK

In total, 23 out of 25 farms sampled presented with *D. gallinae*, through either in person collection or successful collection via cardboard traps, indicating a 92% prevalence of *D. gallinae* across the UK. This is a higher prevalence than observed in both Fiddes et al., (2005) and Guy et al., (2004), where a prevalence of 60-85% prevalence was observed (24, 26). As discussed in the General Introduction (Section 1.4.10), it has been suggested that rises in *D. gallinae* prevalence could be associated with a shift in production systems utilised by the egg laying industry, increasing human population (requiring increased meat and egg production), climate change and regulation changes concerning chemical control However, it should be noted that convenience sampling was employed for this study, and that samples from 17 farms relied upon participant willingness, both of which could result in data bias. The remaining eight samples were provided directly from UK collaborators working on *D. gallinae* (six in total) and as such were already 'collected', two were provided by the British Egg Marketing Board (BEMB, the funding body).

Dermanyssus gallinae was not detected in traps provided by two UK farms (UK24 and UK25). This could be the result of insufficient traps set for detection of *D. gallinae* at these farms or genuine absence of infestation. Questionnaire data revealed no clear indication of a difference at these farms compared to infested farms to signify why *D. gallinae* was not detected. Both UK24 and UK25 utilise a free-range system, with UK25 organic. Flock sizes of 3,600 and 18,000 were recorded with age of hens reported to be 20 and 50/52 weeks of age (both farms having two flocks) implying that flocks had been established for long enough that *D. gallinae* colonisation could have occurred, especially UK25. UK24 used a combination of chemical control

with desiccant dusts, a similar approach to most UK farmers, whilst UK25 used desiccant dusts in combination with predatory mites. UK25 noted that they felt their current control measures were ineffective against *D. gallinae*, implying that *D. gallinae* does reside on this farm, it was just not detected during sampling.

It should be noted that this study was designed primarily for the collection of *D. gallinae* samples spatially and temporally, to aid in genetic analyses later described in this thesis. As a result, there are a lot of limitations to basing prevalence figures from the data sampled, as farms were biased to have an issue with *D. gallinae*. In order to fully understand a more up to date prevalence of *D. gallinae* sampling would need to be conducted over a far greater number of farms, both known to have *D. gallinae* and not currently experiencing issues with greater representation from all production systems.

3.4.2 UK farm overview

From the UK farms sampled, the majority were free-range (88%) with 36% having organic status and 11% of farms used an intensive production system. A flat-deck housing system was the most used housing type and a manure pit was used by almost half of farmers. The average farm age at time of sampling was 24.8 years with 78% of farms having less than 5 units on their farm. Almost all free-range farms provided hens access to wooden or metal structures, with wooden structures being the most frequent. In terms of litter, almost all farmers used some type of shavings or woodchip/wood pellet bedding and over 75% had never changed their choice of litter type. The most common breed of chicken utilised was Hyline and a range of chicken suppliers were used, with Shea Eggs supplying the largest proportion of farms (22%). Chicken mash/meal was identified as the major source of feed and over half of farmers added at least one feed additive to supplement their chickens' diet.

3.4.3 Representation across the United Kingdom

Across the UK, 64% of farms sampled were from England, 16% from Scotland, 16% from Northern Ireland and 4% from Wales. Wales was the least represented country with just one farm replying to the questionnaire and sending *D. gallinae*.

3.4.3.1 Production system representation

According to estimates from DEFRA, in the UK 42% of eggs come from hens kept in enriched cages, 56% from free-range (including 3% organic) and 2% from barn systems (5). In this study, 88% of farms were free-range and intensive production systems represented by just 11%, making them underrepresented. No barn production systems were sampled, although they

represent a low percentage of production systems used in the UK. A skew in production system representation was also observed in other UK studies, such as in Fiddes et al., (2005), where 55% were kept under free-range conditions, 36% kept in barns and 9% kept in cages (5, 26). However, their study incorporated layer chickens, parent birds and fancy fowl breeds, whilst this study focused solely on commercial layer chickens. In contrast, Guy et al., (2004) had an over representation of intensive farms in their study, with 71% of farms utilising a cage system, 4% utilising a barn system and 25% of farms free-range (24).

3.4.4 Factors associated with *D. gallinae*

To gain an understanding of the current opinion of UK farmers surrounding issues relating to *D. gallinae*, a series of questions were included in the questionnaire addressing seasonal fluctuations, markers of a heavy infestation importance and impact of infestation, current control measures, and perceived effectiveness.

3.4.4.1 Seasonal fluctuation

When asked if they noticed seasonal fluctuation in *D. gallinae* almost 75% of farmers reported yes. Research in Sweden looking at the long-term dynamics of *D. gallinae* in relation to control measures in aviary systems demonstrated that populations were significantly denser in summer time compared to winter (193). This is further supported by a study by Magdas et al., focusing on three intensive poultry farms in Romania over all four seasons; spring, summer, autumn and winter. Results showed that infestation prevalence was highest in the summer season (86.6%) and lowest in the winter season (38.3%) (265).

3.4.4.2 Markers of D. gallinae infestation

Visible mites were identified as the most common sign of a heavy infestation (56% of farms), with central areas identified as the most frequent location for them to be found. Similar results from a questionnaire distributed in Portugal by Waap et al., (2019) demonstrated that 45.8% of Portuguese farmers observed *D. gallinae* on their poultry housing, 41.7% observed clusters of mites on furniture and 66.7% of farmers admitted to their staff complaining about *D. gallinae* and itchy skin (29). A much higher rate of *D. gallinae* impacting on staff was reported by Waap et al., (2019) compared to UK questionnaire results, with just one farmer noting that staff can be affected in the UK.

3.4.4.3 Importance and impact of D. gallinae infestation

The majority of farmers (67%) sampled felt that *D. gallinae* is a significant problem with a smaller proportion (28%) reporting it is as moderate and one farmer reporting that it is not a significant

problem. A drop in production was the issue reported most frequently in relation to *D. gallinae* infestation, with farmers also reporting mortality, egg cleanliness, pale or wrinkly eggs, disease and welfare related factors. Similar results were observed from analysis of questionnaire data collected in the UK by Fiddes et al., (2005). In their study, 60% considered that *D. gallinae* is economically very important and an additional 33% reported infestations to be fairly economically important. Out of the 60%, lowered production was the major problem reported by respondents, followed by anaemia in the hens and blood on eggs. Respondents to their questionnaire also reported mortality, feather or vent pecking and a reduced fertility and hatchability (26). Their study was designed to be distributed to both poultry producers and keepers and as no poultry producers were approached in this study, rates of fertility and hatchability were not remarked upon. Guy et al., (2005) also found that the primary concern associated with *D. gallinae* was the impact on egg quality and production, rather than hen mortality or cost of control (24). From a Portuguese study, results showed that 25% of responders related the presence of *D. gallinae* to decreases in egg production, with 16.7% stating that they observed blood spots on eggs (caused by *D. gallinae* crawling onto eggs and being crushed on the conveyer belt) (29). Research has shown that *D. gallinae* outbreaks can result in a drop in egg production regularly of 10-20% (21, 69, 267).

3.4.4.4 Correlation between age of chickens and infestation level

Almost 75% of farmers reported that they notice a correlation between the age of their chicken and the level of *D. gallinae* infestation. One study by Douifi et al., (2019) studying *D. gallinae* infestations in Algeria found that 14% of farms were infested and that flocks older than 40 weeks were the most heavily affected ($P<0.001$) (262). In contrast, investigations by Gunnarsson (2017) of 54 layer flocks in Sweden showed no relationship between bird age, flock age and density of *D. gallinae* (268) and this was further supported by Waap et al., (2019) who found no association (29). However, it should be noted, density was not studied in this study, only presence or absence of *D. gallinae*. They hypothesise that the lack of correlation between *D. gallinae* population density with hen age and flock size could be indicative that several variables contribute to variations in mite population dynamics (29). This has been further supported by research on the prevalence of *D. gallinae* in Korean layer farms that demonstrated 75% of farms sampled had *D. gallinae*, but that there was no significant correlation between flock size and age in the sampled layer farm buildings (261).

3.4.4.5 Control measures against D. gallinae

The predominant control method utilised by farmers studied was desiccant dusts, followed by chemical and hygiene-based methods. Across the world, control of *D. gallinae* is typically achieved through use of synthetic acaricides (33). Due to 36% of farms involved in this study operating organic husbandry systems, the proportion of farms utilising chemical control may be unrepresentative of the UK, as chemical control is not permitted for organic flocks. More than half of the farmers used more than one control measure, and 78% of farmers used pressure washing and disinfectant to clean their poultry units between flocks.

3.4.5 Conclusion

In conclusion, *D. gallinae* was detected on 92% of UK farms sampled, including 23 of 25 farms and representing all four countries. A further 16 countries were sampled from the rest of Europe, including 84 farms. UK farmers have knowledge surrounding *D. gallinae* infestations on their farm, with a drop in production the most common issue raised. A multitude of production systems, housing type, and type of- and combination of control measures were observed to be in place.

4 Genetic Markers for *Dermanyssus gallinae*

4.1 INTRODUCTION

Molecular markers are highly valued in the field of genetics (269-271). Exploiting DNA polymorphisms through molecular marker technologies has impacted positively on animal and plant breeding and genomic research (272). These techniques have aided in understanding the different behaviour and structure of genomes, and revolutionised the characterisation of genetic variation (272). Research using genetic markers was described more than 80 years ago (273, 274), and development of electrophoretic assays (275) and molecular markers (276-287) has greatly improved understanding in biological sciences (288). Over the years, advances in molecular biology have generated many novel molecular markers (270).

4.1.1 Genetic markers

Molecular markers are nucleic acid segments, usually inherited in Mendelian fashion, which are not necessarily affected by environmental factors and do not always encode particular traits (289). Change occurs faster in some nucleic acid segments compared to others (e.g. non-coding DNA vs. coding DNA) and it is useful to use slowly changing markers to compare less related individuals, for example, across different species, and faster changing markers for closely related individuals (289). Thus, different marker types offer a spectrum of usefulness when fingerprinting populations and individuals. A good marker provides a lot of information per assay or per individual, is reproducible, repeatability, cheap to run and has a low error rate (289).

Techniques used to generate molecular markers include restriction fragment length polymorphism (RFLP) (287), random amplified polymorphic DNA (RAPD) (284), amplified fragment length polymorphism (AFLP) (276), inter-retrotransposon amplified polymorphisms (IRAPs) (290), microsatellites (simple sequence repeats (SSR) or short tandem repeats) (278, 291-293), allozymes, single nucleotide polymorphism (SNP) and expressed sequence tag (EST) markers (294-296) and mitochondrial DNA (297, 298).

Despite the economic, medical and/or veterinary importance of many parasitic arthropods, their evolution and population genetics remain poorly understood, partly due to a lack of suitable molecular markers (299). Mitochondrial genomes are informative in the study of population genetics and evolution of vertebrates and invertebrates (300), including parasitic flatworms and nematodes (301, 302). However, studies on parasitic arthropods have commonly focused on sequences of only a small number of mitochondrial genes (299) and like other parasitic arthropods, published research on genetic diversity in *D. gallinae* is predominately focused on these genes, which to date provide the largest source of data (44, 233-235, 237).

4.1.2 Mitochondrial DNA

The evolution and molecular biology of animal mitochondrial DNA (mtDNA) are well-understood (303-305). Across the animal kingdom, the majority of animals have mtDNA that consists of a short, circular molecule containing ~13 protein-coding, intronless genes that play roles in oxidative phosphorylation (i.e. aerobic respiration (306)). Alongside these, animal mtDNA has 22/23 tRNA genes and two rRNA genes which are part of the mitochondrion's translational machinery (306, 307).

The gene content of mtDNA can vary extensively in some groups of organisms (308, 309). Many parasitic organisms have highly reduced mitochondrial genomes, with some species losing mtDNA completely, retaining mitochondrion-related organelles (310, 311) that sometimes retain their aerobic capacity (312). Despite a lot of bilaterians (i.e. animals with bilateral symmetry) having similar mtDNA complements, diversity in genome structures and gene content exist and this diversity expands in non-bilaterian animals (313). As a marker for molecular diversity, mtDNA has been one of the most popular choices over the last three decades, as described below in section 2.1.3 (212).

4.1.2.1 Mitochondrial genomes of parasitic arthropods

Over 700 species of animals are represented by entirely sequenced mitochondrial genomes of which 27 are parasitic arthropods (299). A review by Shao and Barker (2007) indicates that three are Crustacea (shrimps, crabs and kin), ten are Hexapoda (insects and kin) and the remaining 14 come from the subphylum Chelicerata (mites, spiders, ticks and kin) (299). All 27 genomes are circular and contain 37 genes similar to the mitochondrial genomes of animals: 13 encoding proteins, 22 tRNAs and two rRNAs (299). The number and size of non-coding regions is variable but are present in all 27. Variation in the non-coding regions can be seen spanning from less than 100bp long for the small pigeon louse (*Pseudolychia canariensis*) and wallaby louse (*Heterodoxus macropus*) to over 2100bp for the kissing bug (Triatomine bugs) (314-316). In other parasitic arthropods a range from 400-1000bp in the non-coding region is observed, similar to other animals (299).

4.1.3 Advantages of mtDNA markers

The reasons for selecting mtDNA as a molecular marker are well established and offer both experimental and biological advantages (212). Experimentally, amplification of mtDNA targets is simple due to multiple copies being present in most cells (212) and regions of strong sequence conservation across different animals with minimal duplications, short intergenic regions and no

introns (317). It is possible to generate a signal relating to population history over a short time frame using mtDNA in natural populations as it also has highly variable regions with elevated mutation rate (212, 318). Primer design for mtDNA is facilitated by variable regions typically being surrounded by highly conserved ones (212). Biologically, mtDNA possesses properties that make for an appropriate biodiversity marker. Firstly, mtDNA is maternally inherited, so all sites share common genealogy as the entire genome acts as a non-recombining single locus (212, 318, 319). Secondly, mtDNA is considered to be under nearly neutral evolution (320) because mitochondrial genes are involved in metabolic functions (321) and considered less likely to be involved in adaptive processes compared to some other genes. Finally, it is frequently assumed that the evolutionary rate of mtDNA is clock-like: the absence of selective evolutions meaning that only slightly deleterious and neutral mutations accumulate over time so divergence level approximately reflects divergence times (212).

4.1.4 Mitochondrial vs nuclear gene substitution rate comparison

Mitochondrial genes undergo evolution faster than the majority of genes which are encoded by the nuclear genome (322). The synonymous substitution rate of mitochondrial genes is empirically estimated to occur 1.7-3.4 times faster than the most rapidly evolving nuclear genes and 4.5-9 times faster than the average of all nuclear genes studied (323). These estimates are potentially biased due to conserved mitochondrial genes being typically chosen to facilitate primer design (304). Faster evolutionary rate in mitochondrial genes has been linked to stronger constraints occurring in nuclear genomes due to selection for codon usage (323) and a higher rate of transition mutations (322). Other factors thought to play a role in influencing evolutionary rate are nuclear-mitochondrial interactions and thermal adaptation (324). Research has demonstrated positive selective sweeps in the mitochondria including in response to climate-mediated adaptation in humans (325, 326) and high altitude resistance in monkeys and alpacas (327, 328).

Several genes of the mitochondrial genome are increasingly used to assess phylogenetic relationships among animal taxa (see Simon et al., 1994 for a review (304)). Some of the most commonly utilised for phylogenetic purposes include cytochrome c oxidase subunit 1, cytochrome oxidase 2, cytochrome b, ssu (small subunit), lsu (large subunit) ribosomal RNA (rRNA), and the control region (329-331).

4.1.4.1 Studies on acari species mtDNA

A study of 20 mite species belonging to Tenuipalpidae and Tetranychidae demonstrated similarities in the mode of evolution and characteristics of mtDNA to those recognised for

insects, including codon usage, genetic code and base composition. It is thought that this is because these two arthropod classes share ancestral characters (332). Just like insects, mite mitochondrial sequences are A+T rich, on average 75%, however base composition variation occurs between species (332). This has important implications for the construction of phylogenies and requires the development of specific methods that take this variation into account (333). Evolutionary studies on mites have mainly surveyed the mitochondrial cytochrome c oxidase subunit I gene (COI), whereas in studies on ticks the mitochondrial ribosomal 16S has been used more widely (211).

4.1.5 Studies on *Dermanyssus gallinae* mtDNA

As previously discussed in the General Introduction (sections 2.10.1 and 2.10.3) studies on genetic diversity in *D. gallinae* have used the mitochondrial COI gene (233-237), 16S rDNA (18, 236), and the rDNA ITS regions (45, 233, 235-237). Overall, the COI gene has been the most informative marker. An overview of current research can be found in section 2.10.3 and studies have inferred that intranational and international migrations of *D. gallinae* occur within and between countries. Due to the wealth of sequences for the COI gene in *D. gallinae* available on GenBank it was an ideal marker for this study. However, there are limited data available defining COI diversity among mites from UK farms and no direct comparisons between UK and mainland Europe, which prompted the following work.

4.1.6 Mitochondrial Cytochrome c oxidase subunit 1 (COI)

The COI gene encodes a mitochondrial protein that is a key enzyme in the electron transport chain located in the inner mitochondrial membrane (334). In eukaryotic aerobic organisms, it plays a central role in metabolism (334). Despite the availability of other mitochondrial genes to resolve recent divergence, COI likely provides deeper phylogenetic insight compared to alternatives, including cytochrome *b* (335), due to reports of its slower coding sequence divergence compared to other mitochondrial genes (336).

In 2003, Hebert et al. developed a DNA barcoding system for higher eukaryote genomes based on COI sequence diversity (217). Briefly, they created three COI profiles; 1) one for the seven dominant animal phyla, 2) eight profiles of the largest orders of insects, and 3) one to reflect 200 closely allied lepidopteran species. These profiles provided an overview of COI diversity and a basis for identification through determination of sequence congruences between an ‘unknown’ taxon and the taxa included in the specific COI profiles. At ordinal and phylum levels, amino-acid divergences were analysed using Poisson corrected p-values. Their research established that the 5’ region of amino acid diversity provided sufficient reliability to place

species in higher taxonomic categories, ranging from phyla to orders. The evolutionary rate of COI is rapid enough to allow identification of cryptic species and discrimination at species level (337). Focusing on closely allied species of lepidopterans, known for high species diversity and modest molecular evolution rate, they demonstrated discrimination of species based on nucleotide sequence diversity of this same gene region (216, 217). Since establishment as a barcoding marker, COI has been successfully utilised in a multitude of animal taxa (216, 217, 337-343).

4.2 AIMS AND HYPOTHESES

4.2.1 Aim of the study

The main aim of this study was genetic characterisation of *D. gallinae* isolates from Europe, targeting a fragment of the cytochrome c oxidase subunit 1 (COI) gene, to gain insight into population structure and genetic diversity.

4.2.2 Hypotheses

Hypothesis 2: Shared haplotypes exist between countries sharing a land border
Hypothesis 3: Shared haplotypes are seen between the UK and Europe
Hypothesis 4: There is no significant difference in COI diversity between different UK production systems
Hypothesis 5: No deviation from neutrality is observed for COI in *D. gallinae individuals*

4.3 METHODOLOGY

4.3.1 Sample collection and distribution

A subset of the mites described in General Methodology 2.1 were selected for use in this study.

4.3.1.1 Geographical clustering

Countries were arbitrarily grouped into six geographical clusters (Figure 8) based on spatial proximity and climatic factors, with geographic cluster IDs assigned (Tables 9 and 10). Two clusters were assigned to a single country: the UK (cluster 1) and Denmark (cluster 6). The UK formed one cluster due to physical separation from mainland Europe and Denmark was singled out due to climatic differences in comparison to all other countries.

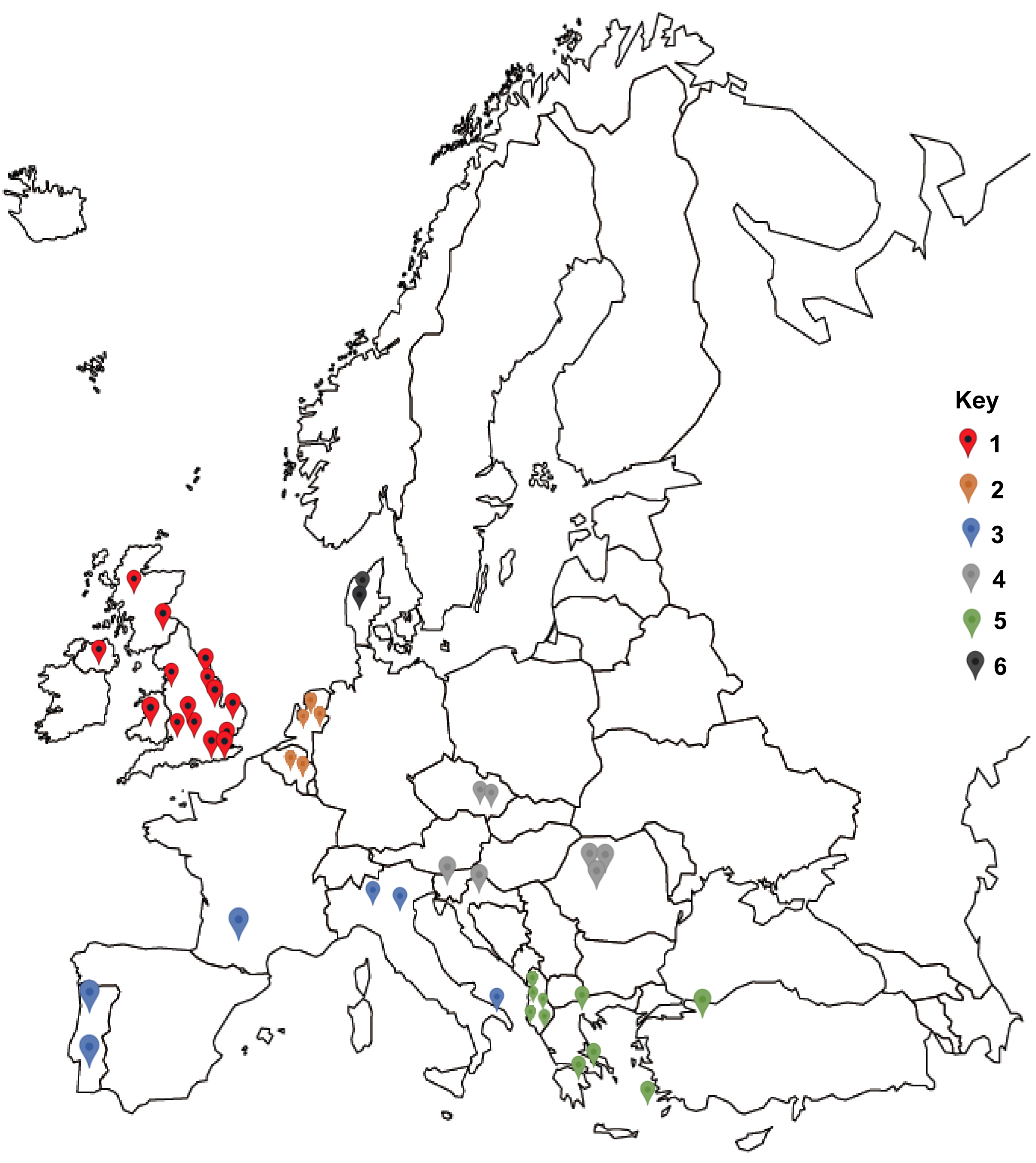


Figure 8: *Map showing the origin of all* D. gallinae *populations analysed in the study, spread across 14 European countries. Six geographical clusters are highlighted in the key (red, orange, blue, green, grey and dark grey pointers). Pointers correspond to approximate locations of closest town or regions for each country outlined in Tables 9 and 10. Geographical clusters are used for some genetic analyses (see below)*

4.3.1.2 United Kingdom

From the UK a total of 15 farms were included, 11 from England, 1 from Wales, 1 from Northern Ireland and 2 from Scotland (Table 9).

Country	County	Sample name(s)	Number of individuals per farm	Production type	Geographic cluster
Wales	Cardiganshire	UK 3.1-3.3	3	Free-range	1
Scotland	Peebleshire	UK 13.1-13.3	3	Intensive	
	Highlands	UK 9.0	1	Free-range	
Northern Ireland	Tyrone	UK 10.1-10.5	5	Free-range	
England	West Sussex	UK 12.1-12.5	5	Free-range	
	Kent	UK 5.0	1	Free-range	
	Gloucestershire	UK 2.1-2.5	5	Free-range	
	Cheshire	UK 4.1-4.4	4	Intensive	
	Durham	UK 1.1-1.2	2	Free-range	
	Oxfordshire	UK 7.0	1	Free-range	
	Shropshire	UK 8.0	1	Intensive	
	Suffolk	UK 11.1-11.3	3	Free-range	
	Lincolnshire	UK 14.1-14.2	2	Free-range	
	Tyne and Wear	UK 15.1-15.2	2	Intensive	
	East Sussex	UK 6.0	1	Free-range	

Table 9: *Location of farms sampled for* D. gallinae *in the UK including the number of individuals, style of production and geographic cluster*

4.3.1.3 Mainland Europe

From mainland Europe, thirteen countries were selected with varying farm numbers, including Albania (5 farms), Belgium (2), Croatia (2), Czech Republic (2), Denmark (2), France (2), Greece (4), Italy (5), Netherlands (5), Portugal (3), Romania (5), Slovenia (3) and Turkey (1). Details about the regions sampled and corresponding number of mites can be seen in Table 10 below and geographic spread can be seen in Figure 8.

Country	No. of isolates per country	Closest Town or Region	Sample names	No. of individuals per region	Geographic cluster
Albania	10	Lushnye	ALB1.1, ALB1.2	2	5
		Berat	ALB2.1, ALB2.2	2	
		Korca	ALB3.1, ALB3.2	2	
		Peshkopi	ALB4.1, ALB4.2	2	
		Durres	ALB5.1, ALB5.3	2	
Belgium	8	Destelbergen	BEL1.1-1.6	5	2
		Evergm	BEL2.1-2.3	3	
Croatia	5	Zagreb	CRO1.1-CRO1.5	5	4
Czech Republic	10	Bohemia	CZH1.1-CZH1.5	5	4
		South Moravia	CZH2.1-CZH2.5	5	
Denmark	9	Vejle	DEN1.1-DEN1.5	4	6
		Jylland	DEN2.1-DEN2.5	5	
France	6	Grenade	FRA1.1-1.6	6	3
Greece	61	Thessaloniki	GRE1.1-GRE1.10	10	5
		Corinth	GRE2.1-GRE2.13	13	
		Leros	GRE3.1-3.25	25	
		Attica	GRE4.1-4.13	13	
Italy	9	Lecce	ITA1.1-1.3	3	3
		Varese	ITA2.1-2.2	2	
		Verona	ITA3.1-3.4	4	
Netherlands	9	Lutten	NET1.1-1.2	2	2
		Barneveld	NET2.1-2.2	2	
		Aalten	NET3.1-3.3	3	
		Unknown	NET4.1-4.2	2	
Portugal	10	Riveria	POR1.1-POR1.4	4	3
		Rego	POR2.1-POR2.6	6	
Romania	7	Tatarlaua	ROM1.1-1.2	2	4
		Cuzdrioara	ROM2.2.2-5	4	
		Floresti	ROM6	1	
Slovenia	7	Tenetiše	SLO1.1-1.3	3	4
		Škofljica	SLO2.1	1	
		Kamnik	SLO3.3-3.5	3	
Turkey	6	Karacaali	TUR1.1-TUR1.6	6	5

Table 10: *Sample locations from Europe (outside of the UK) for* D. gallinae *collection, including the country and number of individual mites sampled*

4.3.2 Polymerase chain reaction (PCR), primer design and amplicon sequencing

A 681bp fragment of the *D. gallinae* COI gene was amplified for individual *D. gallinae* DNA extracts using primers COI1Fyuw114 (5′-AGATCTTTAATTGAAGGGGG-3′) and COI1Ryuw114 (5′-AAGATCAAAGAATCGGTGG-3′) corresponding to nucleotides 61 to 742 (GenBank AM921853; (237)). PCR, agarose gel electrophoresis and Sanger sequencing were carried out as described in the general methodology.

4.3.3 Nucleotide analysis

Population diversity indices including haplotype diversity, number of haplotypes, nucleotide diversity (π) and average number of nucleotide differences (k) were calculated for each country, by geographical clustering and by production system. These tests were done using DnaSP version 6.12.03 (344). Nucleotide diversity was defined as average number of nucleotide differences per site among DNA sequences by pairwise comparison, whilst haplotype diversity (also referred to as gene diversity) was representative of the probability that two randomly sampled alleles will differ (345).

4.3.4 Statistical tests

4.3.4.1 One-way analysis of variance (ANOVA)

To calculate if there was a significant difference between nucleotide diversities and haplotype diversities a one-way analysis of variance (ANOVA) was run on GraphPad Prism 8.4.1.676 with a Tukey's, Sidak's or Dunnett's multiple comparison post-hoc test.

4.3.4.1.1 Post-hoc analysis

A Tukey or Dunnett *post-hoc* test was undertaken after completion of the one-way ANOVA, preferring the Tukey *post-hoc* test (accounting for the possibility of unequal sample sizes) when comparison of every mean with every mean was desired. The Dunnett test was preferred when comparison of every mean to a control mean was desired.

4.3.4.2 Neutrality tests

4.3.4.2.1 Tajima's D

The neutrality test Tajima's D (346) was calculated for the overall dataset, each country and by UK production system. The main principle behind this test is that rapid population expansion associated with a non-neutral process will cause a shift in allele frequency compared with a

neutral Wright-Fisher model, where population expansion follows neutral evolution (347-349). Tajima's D is a statistical measure of the difference between nucleotide polymorphism (θ_w) and nucleotide diversity (π), two estimators of population rate (350).

Under the theory of neutrality, the means of θ_w and π should approximately equal each other, thus the expected Tajima's D value for populations adhering to a standard neutral model will be zero (351). A significant positive or negative deviation relates to a departure from equilibrium neutral expectations, zero, indicating skewed allele frequency distribution (351, 352). Significantly positive Tajima's D values are consistent with balancing selection or population bottlenecks and arise from an excess of intermediate frequency alleles (351, 352). On the other hand, significantly negative Tajima's D values are indicative of positive selection or population expansion through an excess of low frequency alleles (351, 352). Analysis of Tajima's D was completed using DNAsp version 6.12.03 (344).

4.3.4.2.2 Fu and Li's D and F test

The Fu and Li's D and F statistics (353) were calculated for the overall dataset, each country and by UK production system. Fu and Li's D and F tests are similar to Tajima's D as they test for a skewed allele frequency spectrum, however, they make a distinction between recent and old mutations, determined by their position on the branches of genealogies (351, 353). The D and F statistics provide comparison and estimation of population mutation rate based on the number of derived variants which only appear once in a sample (known as singletons) with θ_w and π, respectively (353). Fu and Li's D statistic is based on the comparison of the number of derived singleton mutations with the total number of arrived nucleotide variants, whilst the F statistic is based on a comparison of the number of derived singleton mutations with the mean pairwise difference between sequences (353, 354). As is the case for Tajima's D, the expected values for D and F are zero, with negative and positive deviations providing information about selective or distinct demographic events (351). Analysis of Fu and Li's D and F was completed using DNAsp version 6.12.03 (344).

4.3.5 Phlyogenetic analysis

A total of 195 sequences were aligned and, after trimming of low quality sequence using default parameters in CLC workbench version 8, a 565-bp alignment was used for phylogenetic analysis.

4.3.5.1 Models for estimating distances

Evolutionary distance between paired sequences is typically estimated via the number of nucleotide, or amino acid, substitutions which occur between them (349). Evolutionary distance as a measure is routinely used for inferring phylogenetic trees and estimating the divergence time among individuals, populations, genes and species (355). Most of the commonly used distance estimation methods for amino acid and nucleotide substitutions are incorporated in MEGA-X (356). For the data subset in use, all sites were used for the gaps/missing data treatment option and no branch swap filter was used. From the corresponding output the model with the highest Bayesian Information Criterion (BIC) was selected for phylogenetic tree construction, e.g. for Maximum Likelihood phylogeny, model selection identified the Tamura 3-parameter model.

4.3.5.2 Phylogenetic tree construction

Phylogenetic tree construction was also completed using MEGA-X (356). Maximum Likelihood (ML) and/or Neighbor-Joining (NJ) analysis was completed. Similar parameters were set for both, with an example for ML as follows. A bootstrap method test of phylogeny was assigned with 1000 replications. A nucleotide substitution type was chosen and the corresponding model, as per output from model selection. Rates among sites were selected if required, again following the output from model selection, and if selected five discrete Gamma categories were used. For the data subset in use, all sites were used with the gaps/missing data treatment option. Under tree inference options, standard parameters were used: a nearest-neighbour-interchange ML heuristic method, a default NJ/BioNJ initial tree for ML and no branch swap filter. Bayesian phylogenetic analysis (MrBayes) was determined using TOPALi v2.5 (357). Model selection identified the Hasegawa-Kishino-Yano (HKY) model with gamma distribution (G) and evolutionary invariable (I). Using the HKY+G+I model, the following parameters were used: 2 runs, 5,000,000 generations and 25% Burnin for construction of a MrBayes tree. Interactive Tree of Life (iTOL) version 4 was used for visualisation of MrBayes (358).

4.3.5.3 Network analysis

In parallel, Network 5.0.0.3 (www.fluxus-engineering.com) was used to construct a Median-Joining (MJ) tree (359). Mites with identical sequences were designated as one haplotype. Nodes in each network were colour coded to represent whole countries to provide a visual indication of the relationship of haplotypes within and between countries.

4.3.6 Alignment to published GenBank sequences

Nucleotide sequences generated for this study were aligned to equivalent COI amplicon sequences from Japanese *D. gallinae* individuals produced by Chu et al., (2015) (GenBank accession numbers: LC029457-LC029557) due to utilisation of the same forward and reverse primers. This would enable comparison between European and Asian individualsisolates. Other published COI fragment sequences derived using different primers were uninformative in the absence of sufficient sequence overlap.

4.4 RESULTS

4.4.1 Nucleotide sequence analysis

In total, 195 COI sequences were obtained from mites collected from 82 farms across 14 European countries (GenBank accession no.s LR812284-LR812477). A 565bp alignment representing a fragment of the *D. gallinae* mitochondrial COI gene was analysed. The nucleotide frequencies were 29.05% (A), 40.62% (T), 14.65% (C) and 15.67% (G). The high A-T content (69.67%) observed is consistent when considering the general feature of COI mitochondrial DNA in arthropods. It is similar to COI data produced from research on other mite species, including *Orthithonyssus sylvarium* (235, 347, 360).

4.4.1.1 Nucleotide and haplotype diversity per country compared to the full dataset

Overall, for nucleotide diversity and the average number of nucleotide differences, the lowest scores were observed in Turkey and the highest in Slovenia, whilst the lowest haplotype diversity was seen in Romania and the highest in Belgium (Table 11). Comparison of nucleotide diversity of individual countries compared to the full dataset demonstrated significantly lower diversity for 12 countries (Figure 9) with a difference between values ranging from -0.0007 to 0.02501 (Table 11, Table 54 in Supplementary 10.1.1). Slovenia and the Netherlands showed no significant difference to the full dataset, representing the only two countries to have a similar level of nucleotide diversity in this study (Figure 9).

Five countries showed a significantly lower haplotype diversity compared to the full dataset: Greece, France, Italy, Romania and Turkey (Figure 10) but the remaining nine countries showed no significant difference (Table 55 in Supplementary 10.1.2). Belgium, the Netherlands and the Czech Republic were the only three countries to show higher haplotype diversity, although this was not significant (Figure 10).

Samples	Nucleotide diversity (per site), Pi	Average number of nucleotide differences, k	Haplotype (gene) diversity
All samples	0.02560	14.38598	0.917
UK	0.01403	7.84480	0.901
Greece	0.00419	2.36831	0.521
Albania	0.02124	11.97778	0.889
Belgium	0.01991	11.25000	0.964
Denmark	0.01517	8.556	0.861
Croatia	0.00319	1.8000	0.900
Czech Republic	0.01529	8.62222	0.933
France	0.00153	0.86667	0.733
Italy	0.00345	1.94444	0.722
Portugal	0.02191	12.37778	0.889
Romania	0.00405	2.28571	0.286
Slovenia	0.02630	14.85714	0.857
Turkey	0.00059	0.33333	0.333
Netherlands	0.02557	14.4444	0.944

Table 11: *Nucleotide diversity, average number of nucleotide differences and haplotype diversity for a 565-bp fragment of the COI gene in D. gallinae individuals from individual countries and the full dataset. Generated using DnaSP version 6.*

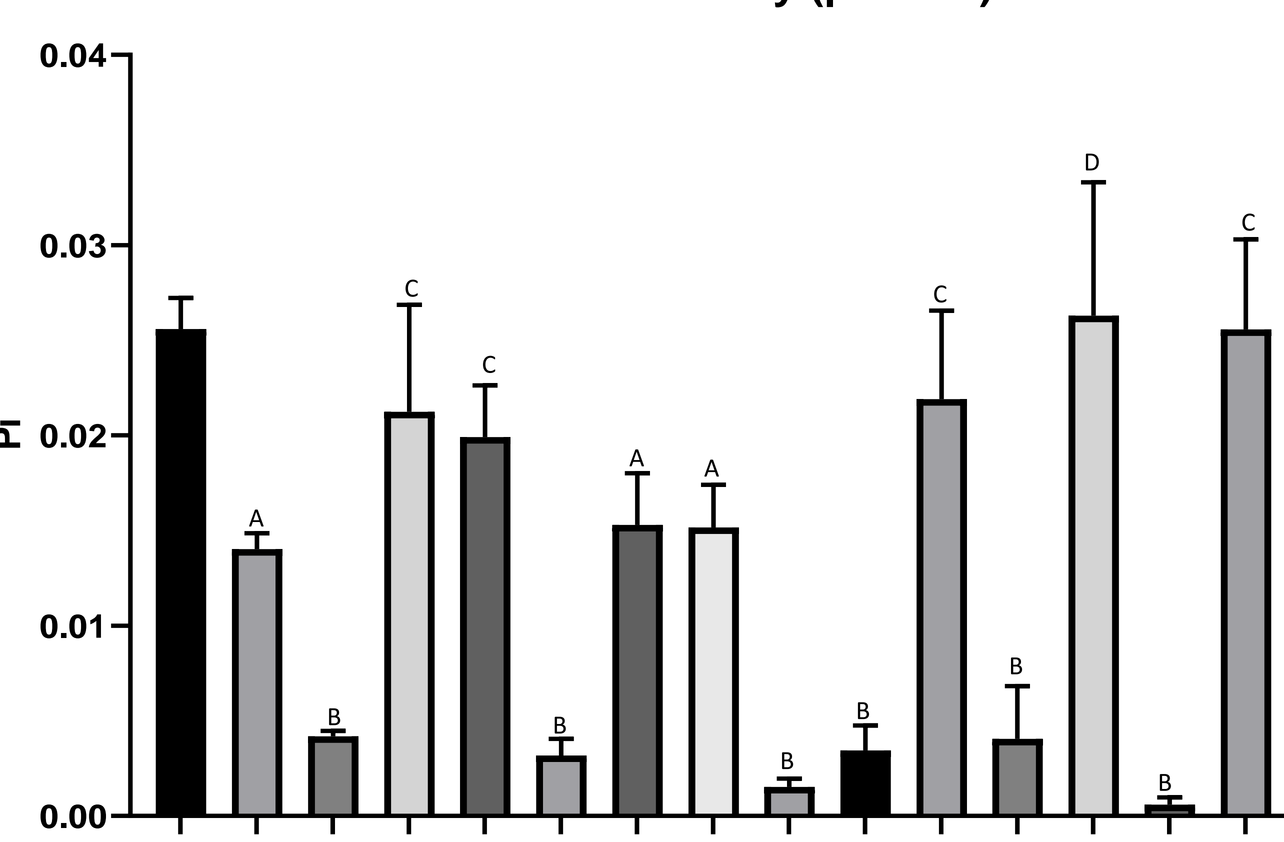


Figure 9: *Nucleotide diversity (Pi) for all samples and individual countries. Comparison of means between countries statistically analysed using a one-way ANOVA. Bars indicated with different letters (ABCD) were found to be significantly different (P<0.05).*

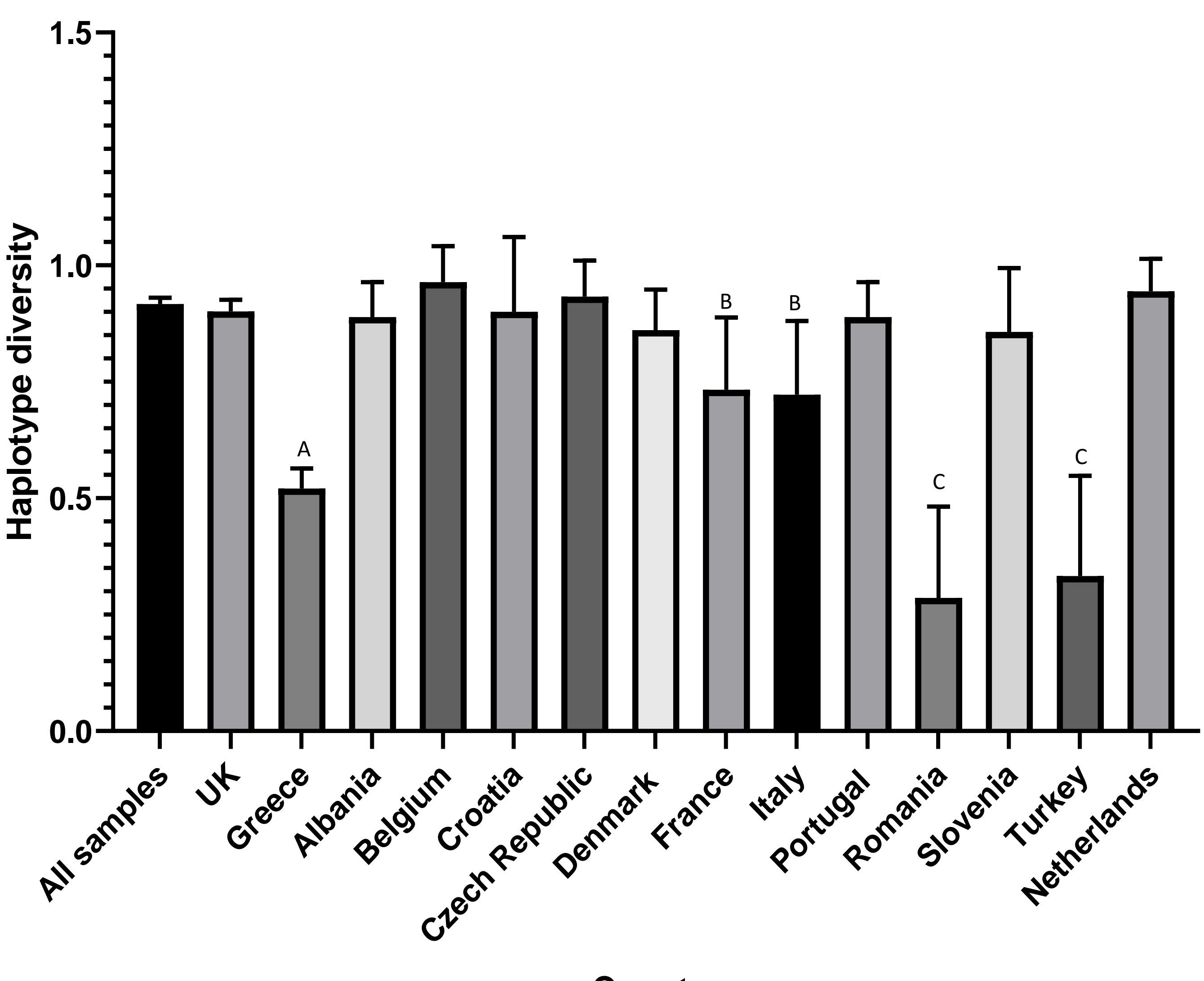


Figure 10: *Haplotype diversity for all samples and individual countries. Statistically analysed using an Ordinary one-way ANOVA with a Dunnett's comparison post-hoc test ran on GraphPad Prism 8.4.3. Standard error bars included Bars indicated with different letters (ABC) were found to be significantly different (P<0.05)*

4.4.1.1.1 Comparison of UK with mainland European countries: Nucleotide diversity

Dermanyssus gallinae collected in the UK were found to have significantly lower nucleotide diversity than the full dataset (Table 58 in Supplementary 10.2.4.1). Comparing UK to the other European countries showed that the UK has a significantly higher nucleotide diversity than six countries: Greece, Croatia, France, Italy, Romania and Turkey (Figure 11). No significant difference was observed between the UK and the Czech Republic or Denmark, implying a similar level of diversity. The UK had a significantly lower diversity compared to five countries: Albania, Belgium, Portugal, Slovenia and the Netherlands (Figure 11).

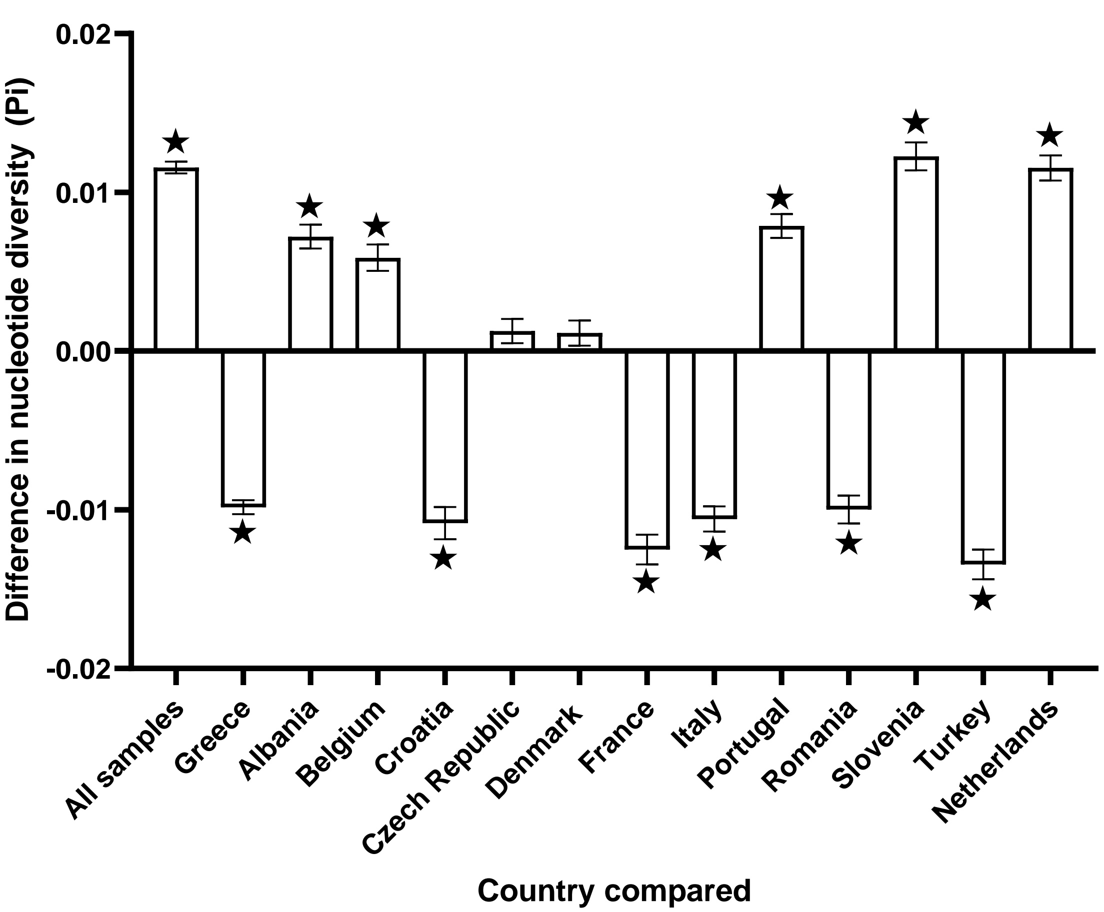


Figure 11: *Comparison of nucleotide diversity in all samples and individual European countries with the UK in a 565-bp fragment COI gene for* D. gallinae. *Statistically analysed using a one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Standard error bars included. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001.*

4.4.1.1.2 Comparison of the UK with mainland European countries: Haplotype diversity

No significant difference was observed between the mean haplotype diversity for the UK and the full dataset (Table 59 in Supplementary 10.1.4.2). A significant difference was only seen between the UK and five countries: Greece, France, Italy, Romania and Turkey, all of which presented significantly lower haplotype diversity (Figure 12). Whilst the UK has lower haplotype diversity than Belgium, the Czech Republic and the Netherlands (Figure 12), this difference was not significant. No significant difference in haplotype diversity was detected when compared to the remaining five countries: Albania, Croatia, Denmark, France, Portugal and Slovenia (Figure 12). This implies a similar level of haplotype diversity amongst the UK and eight of the countries analysed as part of this study.

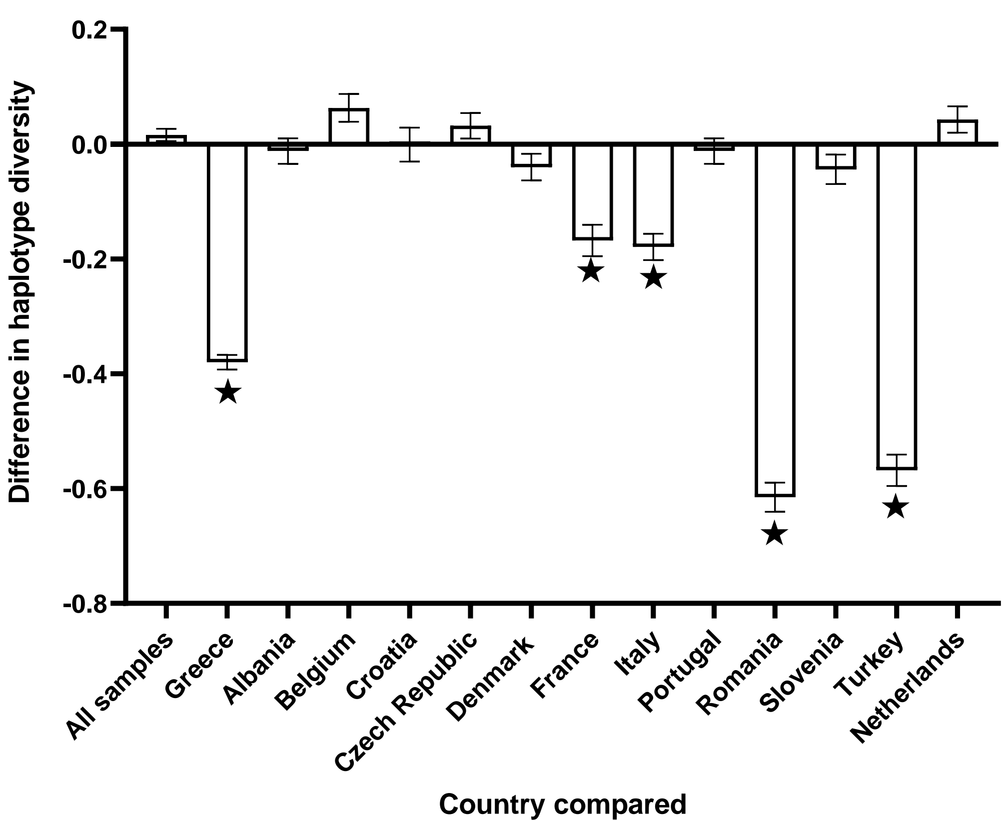


Figure 12: *Comparison of haplotype diversity in all samples and individual European countries to the UK in a 565-bp fragment COI gene for D. gallinae. Statistically analysed using an Ordinary one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Standard error bars included. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001.*

4.4.1.2 Nucleotide and haplotype diversity by geographical cluster

Six geographical clusters were assigned (Tables 9 and 10, Figure 8) based on spatial proximity and climatic factors when comparison of nucleotide and haplotype diversity across full dataset was undertaken (Table 12, Tables 60-61 in Supplementary 10.1.5).

Comparison to the whole dataset revealed that cluster four was the only cluster to have significantly higher nucleotide diversity (Figure 13). The five remaining clusters had significantly lower nucleotide diversity, except for cluster two, where no significant difference to the full dataset was demonstrated. This suggests that cluster two (comprising of Belgium and the Netherlands) has a nucleotide diversity level comparable to the full dataset.

Haplotype diversity was significantly higher for two clusters (two and four) and significantly lower for four clusters (one, three, five and six) (Figure 14) than the whole dataset. The biggest difference in haplotype diversity was observed for cluster five (comprising of Albania, Greece and Turkey) with a difference in mean diversity of -0.222 (Table 12, Figure 14).

Of all six geographic clusters, cluster four (Croatia, Czech Republic, Slovenia and Romania) was the only cluster to have significantly higher nucleotide and haplotype diversity than the full dataset (Figures 13-14). Cluster two showed significantly lower nucleotide diversity but higher haplotype diversity and the remaining four clusters were significantly lower in terms of nucleotide and haplotype diversity (Figures 13-14).

Geographical Cluster	Countries	Nucleotide diversity (per site), Pi	Average number of nucleotide differences, k	Haplotype (gene) diversity
All samples	All	0.02560	14.38598	0.917
Cluster one	UK	0.01403	7.84480	0.901
Cluster two	Belgium and the Netherlands	0.02439	13.77941	0.963
Cluster three	France, Italy and Portugal	0.01297	7.31333	0.877
Cluster four	Croatia, Czech Republic, Slovenia and Romania	0.02924	16.46032	0.955
Cluster five	Albania, Greece and Turkey	0.01182	6.66439	0.695
Cluster six	Denmark	0.01517	0.8556	0.861

Table 12: *Nucleotide and haplotype diversity and the average number of nucleotide differences based on six geographical clusters for a 565-bp fragment COI gene in D. gallinae individuals. Generated on DnaSP version 6.*

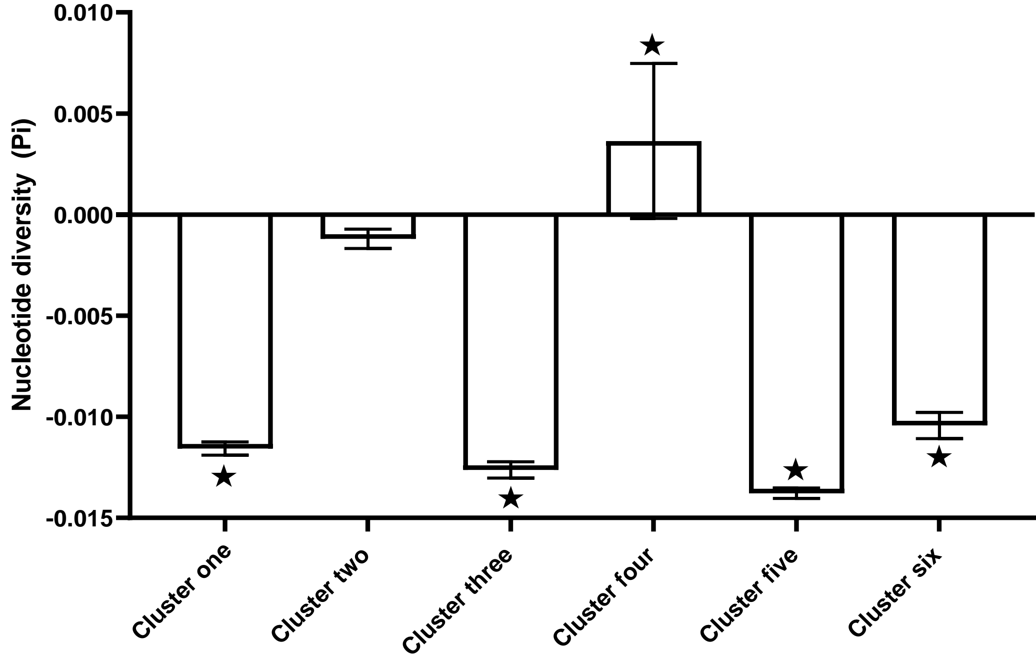


Figure 13: *Difference in nucleotide diversity (Pi) in a 565-bp fragment of the COI gene for D. gallinae from six geographical clusters in comparison to the full data set. Standard error bars shown. Statistically analysed using a y one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Statistical significance indicated by (* *★) representing an adjusted P-value of P<0.0001.*

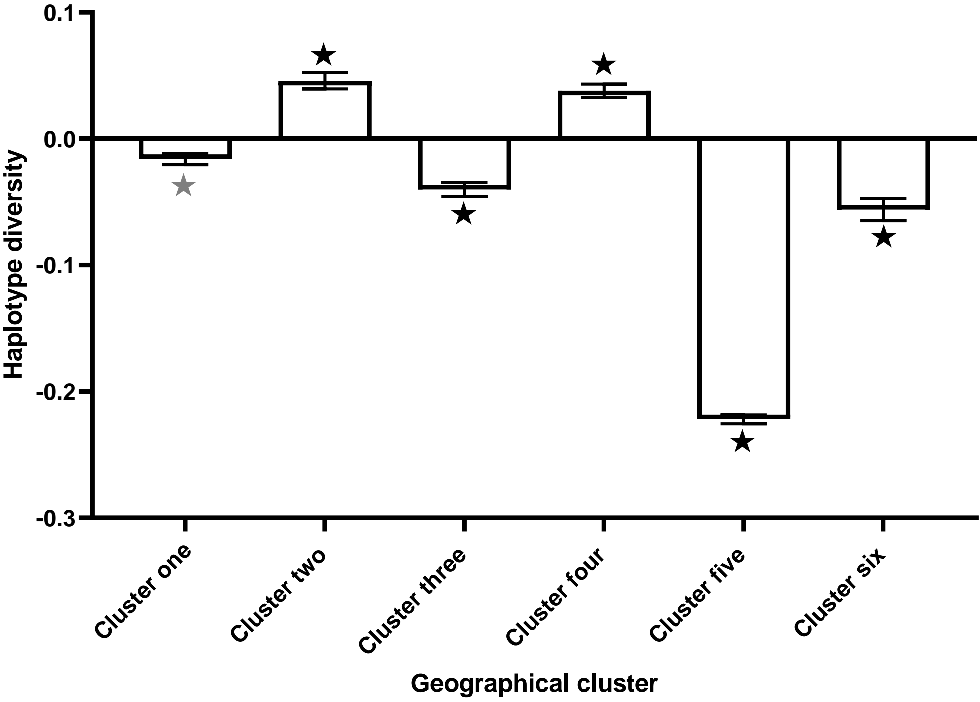


Figure 14: *Difference in mean haplotype diversity in a 565-bp fragment of the COI gene for D. gallinae from six geographical clusters in comparison to the mean haplotype diversity for the full data. Standard error bars shown. Statistically analysed using a one-way ANOVA with a Dunnett's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Statistical significance indicated by (★) representing an adjusted P-value of P<0.0001 or a (★) representing an adjusted P-value of P<0.05.*

4.4.2 United Kingdom and Greece

Two countries were represented by more than 30 individuals; Greece and the UK, permitting more in-depth anaysis. Comparison of nucleotide alignments for each of these countries with the full dataset revealed a significantly lower nucleotide diversity for both the UK and Greece (Figure 9, Table 56 in Supplementary 10.3.3.1). In addition, nucleotide diversity was significantly lower in Greece compared to the UK.

Haplotype diversity was significantly lower for the UK and Greece when compared to the full dataset (Figure 10) with Greece having an observably lower haplotype diversity. Direct comparison of haplotype diversity demonstrated a significantly lower diversity in Greece compared to the UK (Table 57 in Supplementary 10.3.3.2).

4.4.3 Nucleotide and haplotype diversity by production system

Across differing UK production systems, nucleotide diversity was significantly higher for intensive layer systems compared to both the full UK data set and free-range but was significantly lower for haplotype diversity (Table 13, Figure 15, Table 62 in Supplementary 10.1.6.1). Comparison of intensive and free-range haplotype diversity showed free-range had significantly higher diversity (Table 63 in Supplementary 10.1.6.2). Nucleotide diversity was significantly lower for free-range compared to the UK data set, but no difference was seen for haplotype diversity (Table 13, Figure 15).

Production system	Nucleotide diversity (per site), Pi	Average number of nucleotide differences, k	Haplotype (gene) diversity
UK layer	0.01403	7.84480	0.901
Free-range layer	0.01323	7.39409	0.894
Intensive layer	0.01713	9.578	0.844

Table 13: *Nucleotide and haplotype diversity and the average number of nucleotide differences based on production system for a 565-bp fragment COI gene in D. gallinae individuals. Generated on DnaSP version 6.*

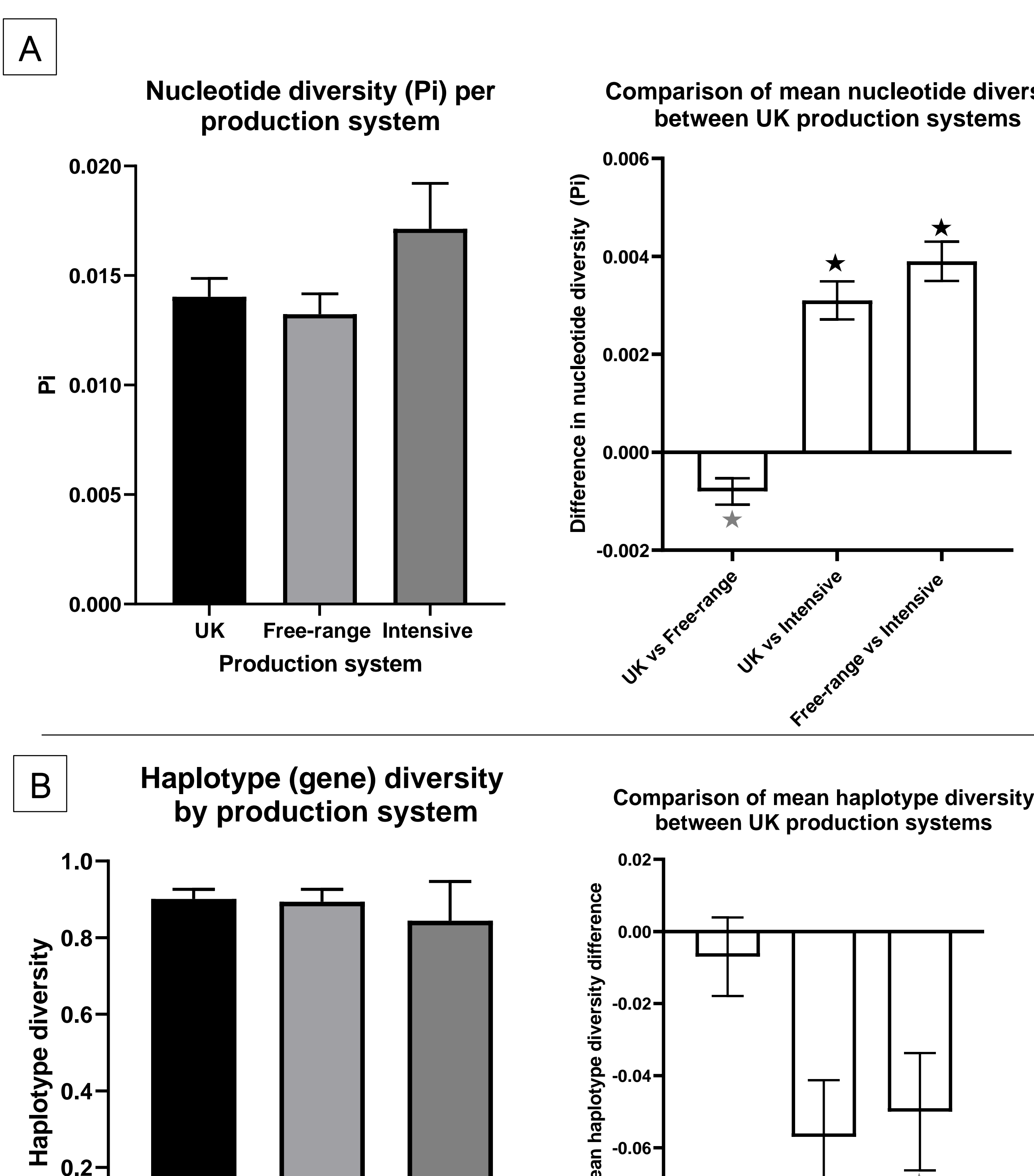


Figure 15: *Graphs depicting the difference in (A) nucleotide diversity and (B) haplotype diversity when comparing different UK production systems Standard error bars shown. Statistically analysed using an Ordinary one-way ANOVA with a Tukey's multiple comparison post-hoc test ran on GraphPad Prism 8.4.3. Statistical significance of P<0.0001 indicated by (★) and P<0.05 by a (★).*

4.4.4 Neutrality tests

4.4.4.1 Tajima's D

The results of Tajima's D test are presented in Table 14, including associated **P**-values calculated by DnaSP. Values ranged from -1.57597 to 2.83215 and analysis of the full dataset and data from 12 of the 14 countries yielded no significant differences from 0. Similarly, assessing sequences from free-range and intensive farms in the UK found no significant difference. This indicates no deviation from 0, demonstrating neutrality in the COI gene for these countries and systems. Greece and Denmark were the exceptions with positive values, indicating that the mite populations in these countries are under balancing selection or undergoing sudden population contraction. They represent geographic (~1300 miles) and climatic extremes amongst the countries tested which suggests there is possible variation in population structures amongst regions. However, it should be noted that Denmark has a small sample size in this study (nine individuals).

Selection	Tajima's D	Statistical significance
Full dataset	-1.29163	NS P>0.10
UK	0.63401	NS P>0.10
UK – Free-range	0.93349	NS P>0.10
UK – Intensive	0.61266	NS P>0.10
Greece	**2.83215**	**P < 0.01****
Albania	0.62179	NS P>0.10
Belgium	1.40631	NS P>0.10
Denmark	**2.19756**	**P < 0.05***
Croatia	-0.41017	NS P>0.10
Czech Republic	-0.61683	NS P>0.10
France	-0.05002	NS P>0.10
Italy	-1.54052	NS P>0.10
Portugal	1.65415	NS P>0.10
Romania	-1.57597	NS P>0.10
Slovenia	-0.99588	NS P>0.10
Turkey	-0.93302	NS P>0.10
Netherlands	0.16746	NS P>0.10

Table 14: *Results of Tajima's D test for the whole dataset and each individual country with associated simulated P value based on a 565-bp fragment COI gene in D. gallinae individuals. Data generated on DnaSP version 6.*

4.4.4.2 *Fu and Li's F and D Tests*

The results of Fu and Li's D and F tests are presented in Table 15, with associated P-values as calculated by DnaSP. Values for the D statistic ranged from -6.1785 to 1.58335 and for the F statistic from -4.6138 to 1.91667 (DnaSP version 5). Overall, the full dataset was found to differ significantly ($P<0.02$) with negative values for both the D and F statistics (Table 15). A total of 11 countries including the UK showed no significant difference for either D or F statistics and there was no significant difference for the UK when free-range and intensive production systems where tested separately. Denmark and Portugal were found to differ significantly, with positive values for D and F statistics. Greece was found to have a significantly positive F statistic but was not statistically significant for the D statistic.

Selection	Analysis using only biallelic positions				Results from DnaSP V5			
	No. of segregating sites	D	F	Achaz Y	D statistic	P value	F statistic	P value
Full dataset	**104**	**-6.84166**	**-4.92557**	**0.71625**	**-6.17854**	**P < 0.02**	**-4.61377**	**P < 0.02**
UK	26	0.08595	0.39362	1.05018	-0.31209	NS P > 0.10	0.01040	NS P > 0.10
UK – Free-range	23	-0.25587	0.13377	1.51854	-0.25587	NS P > 0.10	0.14444	NS P > 0.10
UK – Intensive	24	0.07126	0.21137	1.19196	0.07126	NS P > 0.10	0.23248	NS P > 0.10
Greece	5	1.07924	1.82143	2.32493	**1.07924**	**NS P > 0.10**	**1.91667**	**P < 0.05**
Albania	30	0.91853	0.86620	-0.25128	0.91853	NS P > 0.10	0.95301	NS P > 0.10
Belgium	23	1.28378	1.33139	0.43242	1.28378	NS P > 0.10	1.46075	NS P > 0.10
Denmark	16	1.54221	1.73396	1.25988	**1.54221**	**P < 0.02**	**1.90378**	**P < 0.02**
Croatia	4	-0.41017	0.41017	N/A	-0.41017	NS P > 0.10	-0.41751	NS P > 0.10
Czech Republic	28	-0.86781	-0.82587	0.39313	-0.86781	NS P > 0.10	-0.90855	NS P > 0.10
France	2	0.06221	0.03847	-0.66667	0.06221	NS P > 0.10	0.03984	NS P > 0.10
Italy	8	-1.59647	-1.61311	-1.15857	-1.59647	NS P > 0.10	-1.76646	NS P > 0.10
Portugal	26	1.58335	1.64192	0.42948	**1.58335**	**P < 0.02**	**1.80610**	**P < 0.02**
Romania	8	-1.66615	-1.65997	N/A	-1.66615	NS P 0.10 > P > 0.05	-1.79992	NS P 0.10 > P > 0.05
Slovenia	44	-0.96068	-0.97732	-0.69667	-0.96068	NS P > 0.10	-1.07161	NS P > 0.10
Turkey	1	-0.95015	-0.94938	N/A	-0.95015	NS P > 0.10	-0.96473	NS P > 0.10
Netherlands	38	0.38359	0.33794	-0.35667	0.38359	NS P > 0.10	0.37176	NS P > 0.10

Table 15: *Fu and Li's D and F test results with associated P-values for the full dataset, UK (including separate production systems) and individual European countries based on a 565-bp fragment COI gene in D. gallinae individuals. Data generated on DnaSP version 6.*

4.4.4.3 Neutrality tests: Full dataset sliding window comparison

Analysis of the full dataset produced a minus value for Tajima's D as well as Fu's D and F statistics, but only Fu's D and F statistics were significant (Tables 14-15). Significant D and F statistics were found at 14 of the 20 midpoints on the sliding window, 12 with P<0.02 and two with P<0.05 (Table 64 in Supplementary 10.1.7, Figure 16). No significance was found for D and F statistics between midpoints 302-452, corresponding to nucleotide positions 253-502 (Table 64 in Supplementary 10.1.7). None of the midpoints were found to be significant for Tajima's D (Table 64 in Supplementary 10.1.7, Figure 16). Plotted on a sliding window graph, Tajima's D followed a trend most like nucleotide diversity, whilst D and F statistics followed a trend similar to each other (Figure 16).

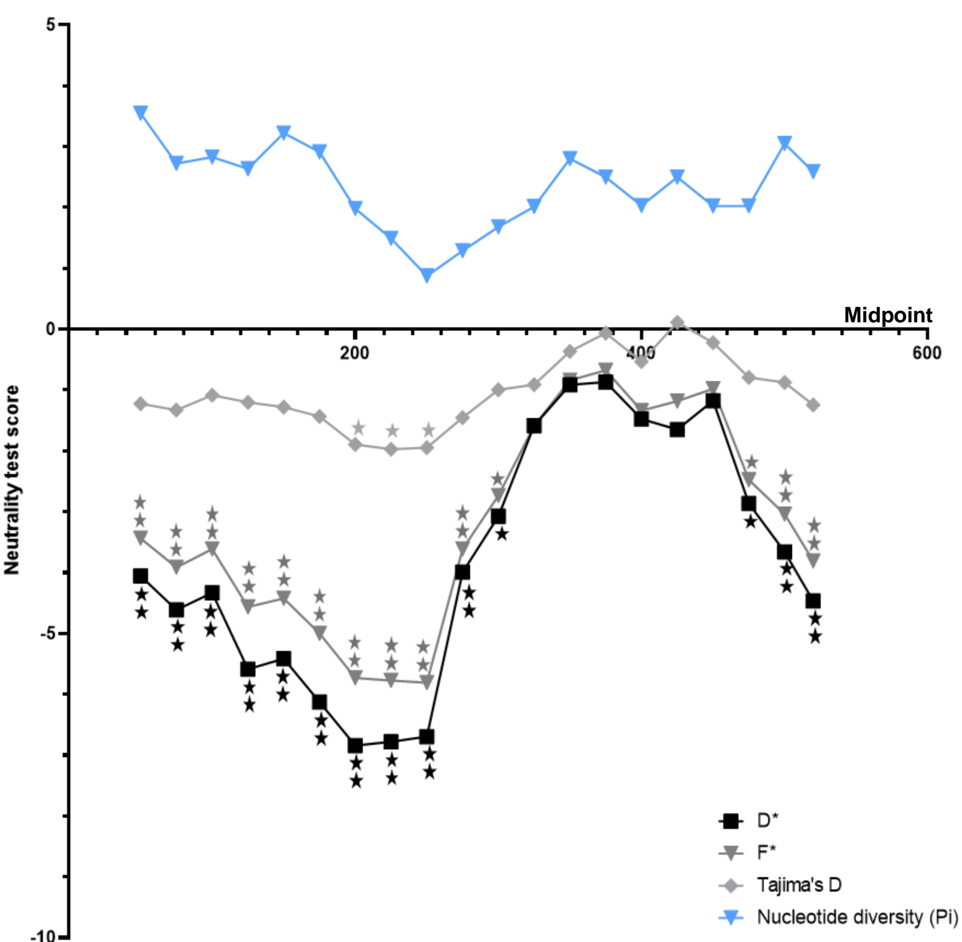


Figure 16*: Sliding window comparison of Fu and Li's D and F tests with Tajima's D and nucleotide diversity for the full dataset. Sliding window computed on DnaSP from site 1 to 564, with a sliding window length of 100 sites and step size 25. Significance for Fu and Li's D and F: #, P<0.10; *, P<0.05; **, P<0.02 and for Tajima's D: # P<0.10; ★ P<0.05; ★★P<0.01; ★★★ P<0.001.*

4.4.4.4 Neutrality tests: Greece

Analysis of Greek individuals revealed a significantly positive value for Tajima's D and Fu and Li's F statistic but no significance for D statistic (Tables 14-15). Sliding window computation showed significant values for 1 midpoint (P<0.10) for F statistic, zero midpoints for F statistic and 4 midpoints for Tajima's D (P<0.05) (Table 65 in Supplementary 10.1.8). No significant midpoints were identified for the first 15 midpoints, corresponding to nucleotide positions up to 376 (Table 65 in Supplementary 10.1.8**Table 67**, Figure 17). Plotted on a sliding graph, all neutrality tests followed a similar trend to nucleotide diversity, including all scoring 0 at midpoint positions 100-250, corresponding to nucleotide positions 51-300 and 374 and 400, corresponding to nucleotide positions 326-450 (Table 65 in Supplementary 10.1.8, Figure 17).

Figure 17*: Sliding window comparison of Fu and Li's D and F test with Tajima's D and nucleotide diversity for Greece. Sliding window computed on DnaSP from site 1 to 564, with a sliding window length of 100 sites and step size 25. Significance for Fu and Li's D and F: #, P<0.10; ★, P<0.05; ★★, P<0.02 and for Tajima's D: # P<0.10; ★ P<0.05; ★★ P<0.01; ★★★ P<0.001*

4.4.5 COI sequence variation in the United Kingdom

A total of 39 COI sequences were obtained from individual mites from the UK, representing 15 farms. Alignment revealed 27 mutations between samples when compared to the consensus (Table 16). Of these eight represented a single farm, seven an individual country, and six were detected in a single mite individual (Table 16). No insertions or deletions were seen. Of the 15 farms sampled, five were represented by a single mite and were disregarded from intra-farm analysis. From the remaining ten farms, six demonstrated intra-farm variation and four showed no intra-farm variation (including Northern Ireland). Twenty five of the 27 mutations found to be present in the UK (Table 17) originated from at least one farm that hosted intra-farm variation (Table 17). At seven sequence positions, only one of the five farms showed variation, with three of these from a single farm (UK15).

Base pair position	Consensus	Mutation	No. of individuals consensus	No. of individuals with mutation	Mutation found from a single country
9	A	G	34	5*	Northern Ireland
33	T	C	38	1*	Scotland
36	C	T	24	15	-
37	T	C	25	14	-
60	T	C/A	37	1/1*	Wales/England
69	A	G	24	15	-
123	A	G	36	3	-
126	A	G	24	15	-
154	T	C	37	2	-
162	T	A	24	15	-
167	C	T	38	1*	England
174	A	G	36	3	-
189	C	T	21	18	-
300	T	C	34	5	-
336	T	C	24	15	-
360	A	G	28	11	-
396	T	C	24	15	-
411	C	T	24	15	-
450	G	A	26	13	-
456	T	C	34	5	-
465	C	T	38	1*	England
480	A	T	38	1*	England
498	T	C	37	2	-
528	T	C	21	18	-
534	A	G	27	12	-
546	T	C	38	1*	England
549	G	A	38	1*	England

Table 16: *Variable positions for a 565-bp fragment of the COI gene for UK D. gallinae individuals when compared to the consensus sequence. Information regarding the base pair at the consensus and the mutation(s) present with the number of individual provided. Mutations which were found in a single country from the UK are indicated in the far-right column, with any mutations present in multiple indicated by a dash (-). A * is used to indicate that all individuals belong to a single farm.*

Base pair position	Farm	Country	Consensus	Mutation	No. of individuals consensus	No. of individuals with mutation	Total no. of individuals
33*	UK13	Scotland	T	C	1	2	3
36	UK14	England	C	T	1	1	2
	UK3	Wales			1	2	3
37	UK2	England	T	C	2	3	5
	UK3	Wales			1	2	3
60	UK3	Wales	T	C	1	2	3
	UK15	England	T	A	1	1	2
69	UK3	Wales	G	A	1	2	3
	UK14	England			1	1	2
123	UK13	Scotland	A	G	1	2	3
	UK14	England			1	1	2
	UK15	England			1	1	2
126	UK3	Wales	A	G	1	2	3
	UK14	England			1	1	2
153	UK13	Scotland	T	C	1	2	3
	UK14	England			1	1	2
162	UK3	Wales	T	A	1	2	3
	UK14	England			1	1	2
167*	UK11	England	C	T	1	2	3
174	UK11	England	A	G	1	2	3
	UK13	Scotland			2	1	3
189	UK3	Wales	T	C	2	1	3
	UK14	England			1	1	2
300	UK3	Wales	T	C	1	2	3
	UK14	England			1	1	2
336	UK3	Wales	T	C	1	2	3
	UK14	England			1	1	2
360	UK3	Wales	A	G	1	2	3
	UK11	England			2	1	3
	UK15	England			1	1	2
396	UK3	Wales	T	C	2	1	3
	UK14	England			1	1	2
411	UK3	Wales	C	T	2	1	3
	UK14	England			1	1	2
450*	UK2	England	G	A	3	2	5
456	UK3	Wales	T	C	2	1	3
	UK14	England			1	1	2
465*	UK15	England	C	T	1	1	2
480*	UK15	England	A	T	1	1	2
498	Uk13	Scotland	T	C	2	1	3
	UK14	England			1	1	2
528	UK3	Wales	T	C	2	1	3
	UK14	England			1	1	2
546*	UK11	England	T	C	2	1	3
549*	UK15	England	G	A	1	1	2

Table 17*: Intra-farm variation observed in D. gallinae individuals for a 565-bp fragment of the COI gene from UK farms. Base pair position, farm, country and information regarding number of samples with the mutation compared to the consensus sequence*

4.4.5.1 *Intra-farm variation by production system*

Out of the twenty five nucleotide positions with intra-farm variation present, five were shared by both production systems, five solely from intensive farms and the remaining fifteen related to free-range farms (Table 18, Figure 18).

Base pair position	Free range	Intensive	Both
33		Y	
36	Y		
37	Y		
60		Y	
69	Y		
123			Y
126			Y
153			Y
162	Y		
167	Y		
174			Y
189	Y		
300	Y		
336	Y		
360			Y
396	Y		
411	Y		
450	Y		
456	Y		
465		Y	
480		Y	
498	Y		
528	Y		
546	Y		
549		Y	

Table 18: *Intra-farm variation observed in D. gallinae individuals for a 565-bp fragment of the COI gene for different UK production systems. Base pair position relating to mutations detailed in Table 17 with columns indicating presence in free-range farms, intensive farms or in both production systems.*

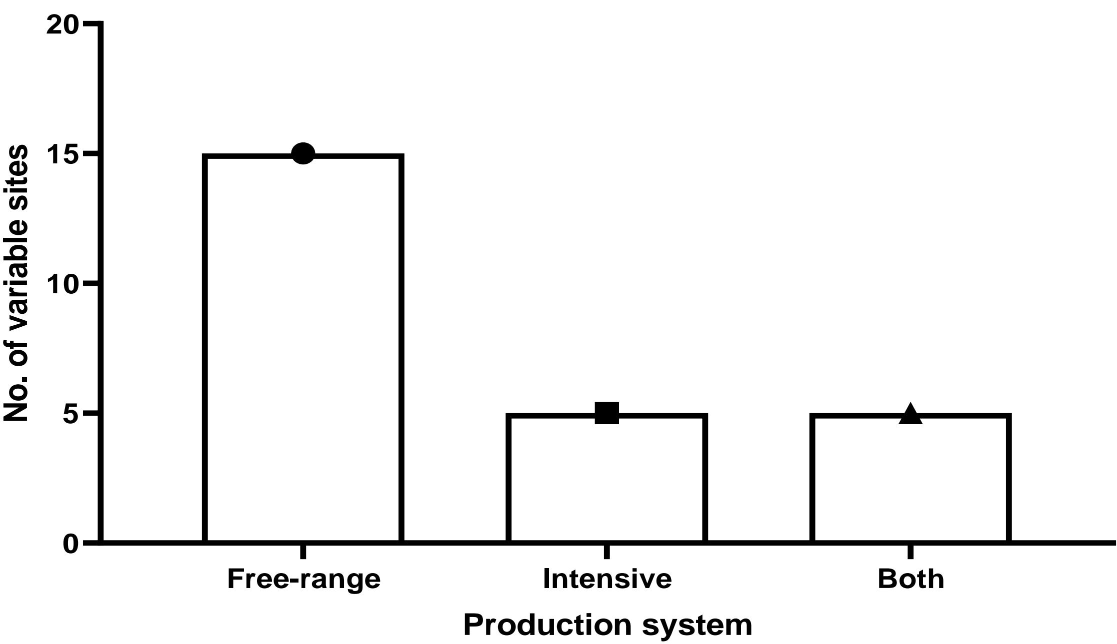


Figure 18: *The number of variable sites from a 565-bp fragment of the COI gene in D. gallinae individuals demonstrating intra-farm variation attributed to UK production systems (free-range (circle), intensive (square), or both(triangle))*

4.4.6 Intra-farm variation: Greece

Intra-farm variation was detected in mites from all four Greek farms at five nucleotide positions at variable rates, with between 20-68% of samples from a single farm presenting the mutation in comparison to the consensus (Table 19).

Base pair position	Farm	Consensus	Mutation	No. of individuals consensus	No. of individuals with mutation	% of individuals with mutation	Total no. of individuals
42	THE	T	A	3	7	70%	10
	LER			8	17	68%	25
	ATT			6	7	54%	13
	COR			7	6	46%	13
305	THE	T	C	8	2	20%	10
	LER			18	7	28%	25
	ATT			9	4	31%	13
	COR			7	6	46%	13
455	THE	A	G	8	2	20%	10
	LER			17	8	32%	25
	ATT			8	5	38%	13
	COR			6	7	54%	13
461	THE	T	C	10	2	20%	10
	LER			17	8	32%	25
	ATT			7	6	46%	13
	COR			6	7	54%	13
539	THE	A	G	3	7	70%	10
	LER			17	8	32%	25
	ATT			6	7	54%	13
	COR			7	6	46%	13

***Table 19**: Base-pair positions from a 565bp fragment of the COI gene from D. gallinae individuals collected from Greek farms displaying Intra-farm variation, including information regarding the number of individuals and percentage of individuals with the mutation*

4.4.7 Phylogenetic analysis of a 565-bp COI fragment in *D. gallinae individuals from* the United Kingdom

Phylogenetic analysis of the 39 COI sequences from the UK revealed two major haplogroups, with a total of seventeen haplotypes (Figure 19). Ten haplotypes belonging in haplogroup 1 and seven haplotypes in haplogroup 2. At a country level, Northern Ireland grouped into one haplotype (haplogroup 1), which was not shared with England, Scotland or Wales, although all individuals came from a single Northern Irish farm. Individuals from Scotland, England and Wales were found distributed in both haplogroups but only one haplotype shared individuals from all three countries.

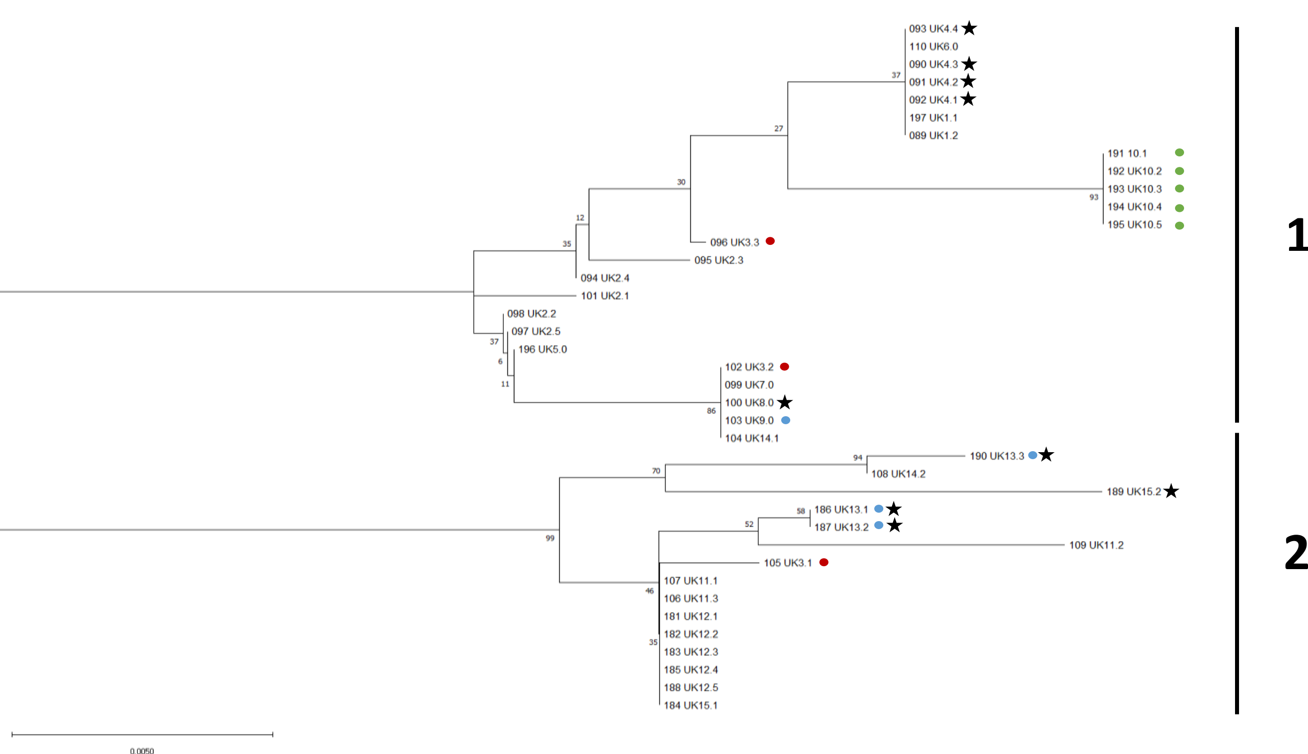


Figure 19: *Phylogenetic tree of partial COI sequences representing* D. gallinae *individuals collected in the UK, inferred using the Tamura 3-parameter and maximum-likelihood (361). A discrete Gamma distribution was utilised to model evolutionary differences among sites (5 Categories (+G, parameter = 0.0500)). A total of 559 positions were used in the analysis, encoding 39 nucleotide sequences. All evolutionary analysis was completed with MEGA X (356). Countries from the UK are indicated as follows: England = no colour, Red = Wales, blue = Scotland, green = Northern Ireland. To differentiate production systems, intensive farms are indicated by a black star.*

4.4.8 Phylogenetic analysis of a 565-bp COI fragment in *D. gallinae* individuals from European and UK individuals

Maximum-likelihood phylogenetic analysis of the 195 COI sequences obtained revealed 76 distinct haplotypes that clustered into three main haplogroups: A, B and C (Figure 20). Group A consisted of 22 haplotypes from 10 countries, group B 34 haplotypes from seven countries and group C 20 haplotypes from seven countries. Haplogroup 1 described from UK mite samples in section 4.4.5 were located in main haplogroup A, while UK haplogroup 2 was located in main haplogroup B. The three major haplogroups diverged into a further six sub-lineages designated as Aa, Ab, Ba, Bb, Ca, Cb (Figure 20). Group Aa included 14 haplotypes, group Ab nine haplotypes, group Ba four haplotypes, group Bb consisted of 30 haplotypes, group Ca four haplotypes and Cb consisted of 16 haplotypes.

In total, sequences from 8 out of 14 countries clustered into a single haplogroup, 4 out of 14 countries into two haplogroups and 2 out of 14 countries into three haplogroups. Albania and the Netherlands were the only two countries where *D. gallinae* individuals were represented by sequences from all three haplogroups. Indiviuals from Greece and Romania were only found in sub-groups Aa and Ab and Turkish Individuals were only found in sub-group Ab (two haplotypes), but it should be noted that only one farm from Turkey was sampled. Denmark was the only country to be found solely in sub-groups Ba and Bb, representing three out of the four haplotypes found in sub-group Ba. Sub-haplogroup Ca was the only subgroup to represent a single country, entirely consisting of six *D. gallinae* individuals collected across three farms from Portugal. The remaining four Portugese individuals were clustered into subgroup Aa (three) and Cb (1). Bayesian phylogenetic analysis supported topology from ML with three main haplogroups: A, B and C that diverge into six subgroups: Aa, Ab, Ba, Bb, Ca, Cb (Figure 20). Variation in individual haplotypes within subgroups was observed when comparing ML and MrBayes trees (Figure 20) but overall tree topology remained consistent. Identical clustering of countries in haplogroups was observed, eight clustering in a single haplogroup, four in two haplogroups and two in three haplogroups (Albania and the Netherlands).

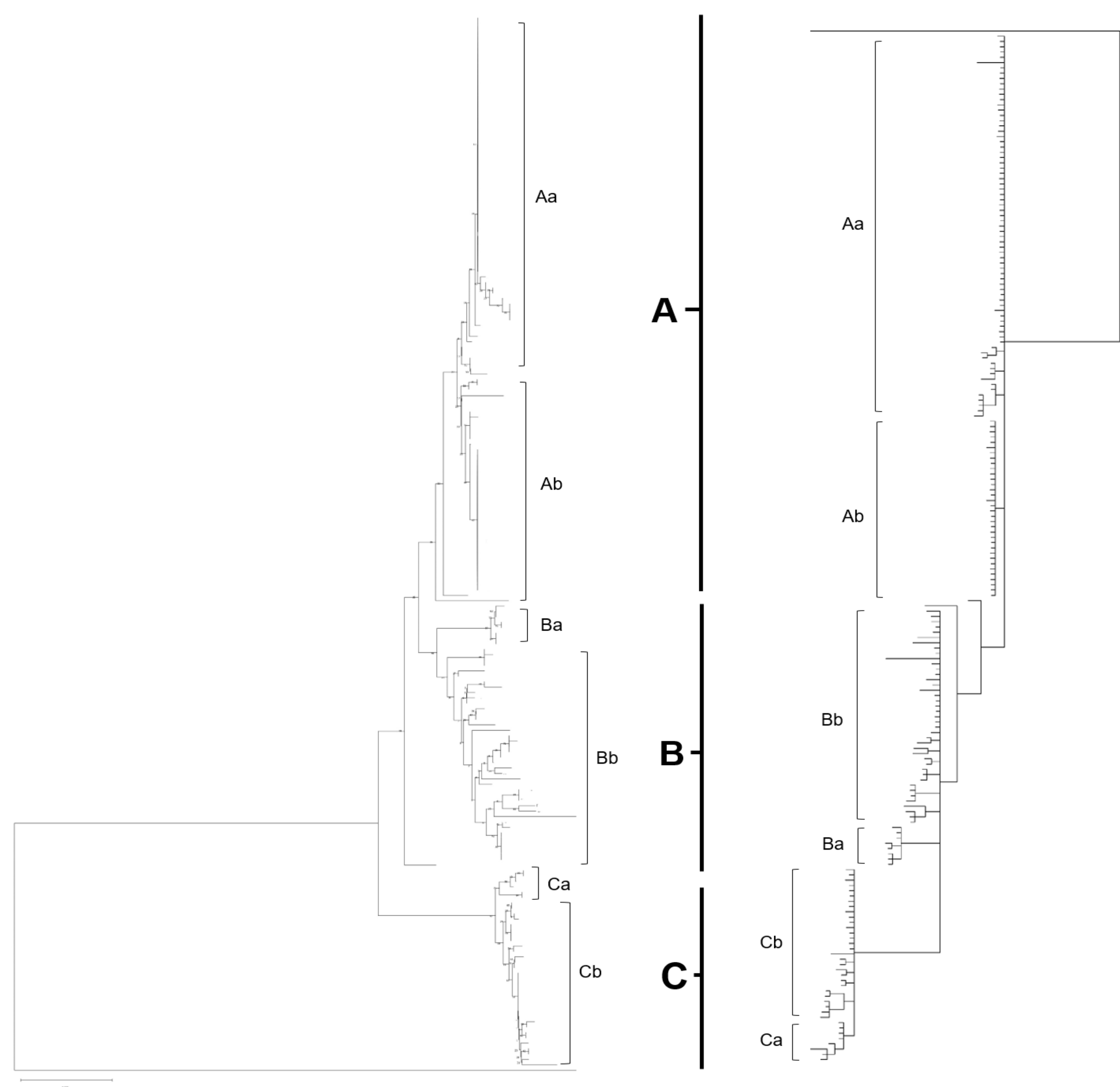


Figure 20: *(A) Phylogenetic tree of all European and UK individuals sequenced as part of this study. Inferred using the Tamura 3-parameter and maximum-likelihood with 1000 replicates (Tamura, 1992). A gamma distribution was utilised to model evolutionary differences (shape parameter = 0.5). A total of 565 positions were used in the analysis, encoding 196 nucleotide sequences. All evolutionary analysis was completed with MEGA X (356).(B) Bayesian phylogenetic tree of all European and UK individuals sequenced as part of this study. Inferred using the HKY+G+I model with 2 runs, 5,000,000 generations and 25% Burnin. A total of 565 positions were used in the analysis, encoding 196 nucleotide sequences. Evolutionary analysis completed on TOPALi (357) and edited on iTOL (358).*

4.4.9 Network analysis

The geographic origin of sequences placed within each of the haplogroups identified in the phylogenetic tree can be observed in Figure 21. Turkey and Romania were the only countries represented by just two haplotypes. Five farms from Romania were sampled, all located in haplogroup A, with four farms clustered in one haplotype and the remaining farm an orphan haplotype.

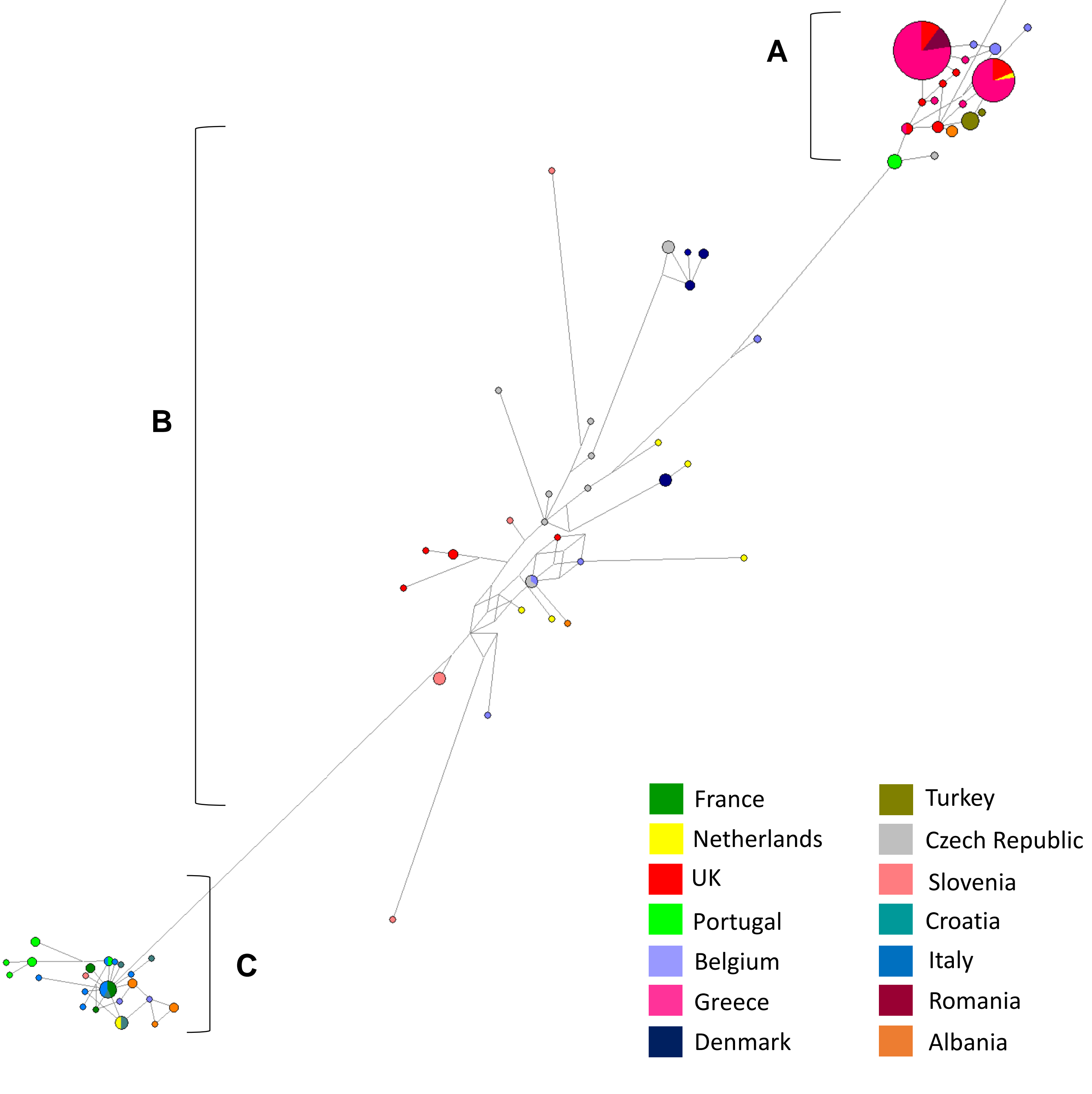


Figure 21: *Network analysis of all European and UK D. gallinae individuals sequenced in the study with the three main haplogroups labelled, A, B and C. Median-joining tree analysis was completed on Network 5.0.0.3. Countries are colour coded in the key in the bottom right. A total of 554 positions were used in this analysis, encoding 195 nucleotide sequences.*

4.4.10 Comparative analysis with sequences published in GenBank

Network analysis comparing Japanese and UK individuals showed three main haplogroups (Figure 22). One consisted purely of Japanese samples (haplogroup 3), including one dominant haplotype, and two further haplogroups contained a mixture of Japanese and UK haplotypes. Haplogroups 1 and 2 refer to the same phylogenetic groupings of UK samples as seen in Figure 19. England was the only country found to directly share haplotypes with Japan. A total of three shared haplotypes were seen, two made up mostly by Japanese individuals and one more common to English individuals. No shared haplotypes were observed between Japan, Wales, Northern Ireland or Scotland, although all five countries were found clustered in haplogroup two.

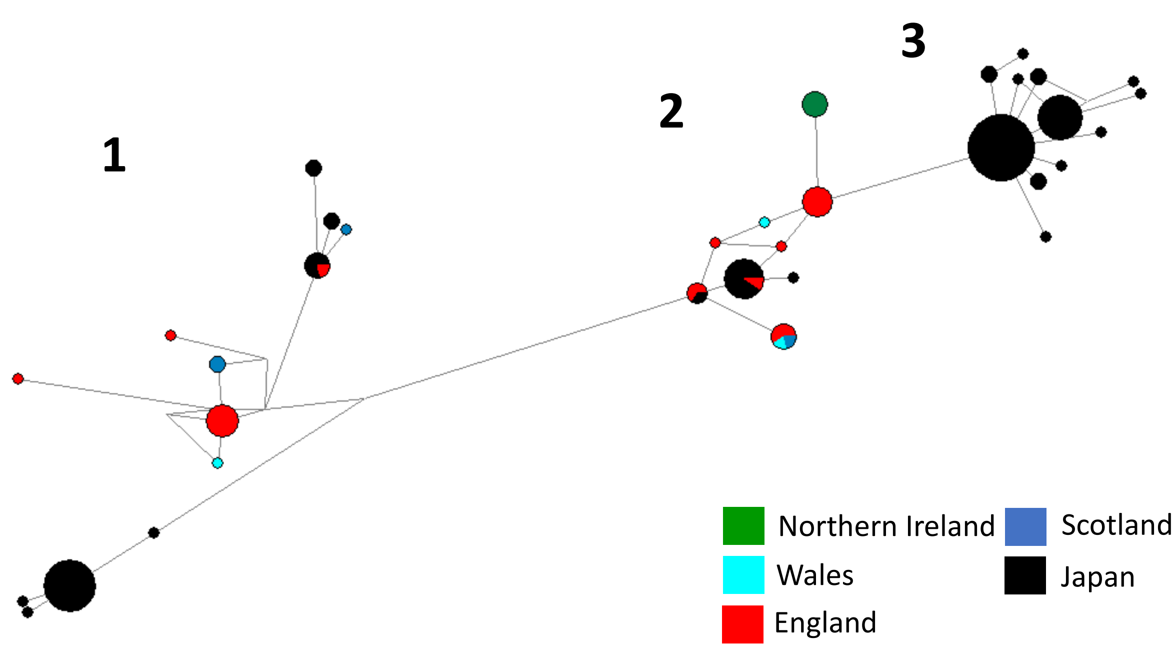


Figure 22: *Network analysis of* D. gallinae *individuals sequenced from the UK and Genbank sequences available from Japan (237). Countries are colour coded in the key in the bottom right. Median-joining tree analysis was completed on Network 5.0.0.3. A total of 554 positions were used in this analysis, encoding 139 nucleotide sequences*

Network analysis confirmed that European and Japanese samples were genetically related, as previously demonstrated (Chu et al., 2015). One haplotype was common to Japan, the UK and Greece in haplogroup A (Figure 16). In haplogroup B, another haplotype was common to Japan, Belgium, the Czech Republic and the UK (Figure 23).

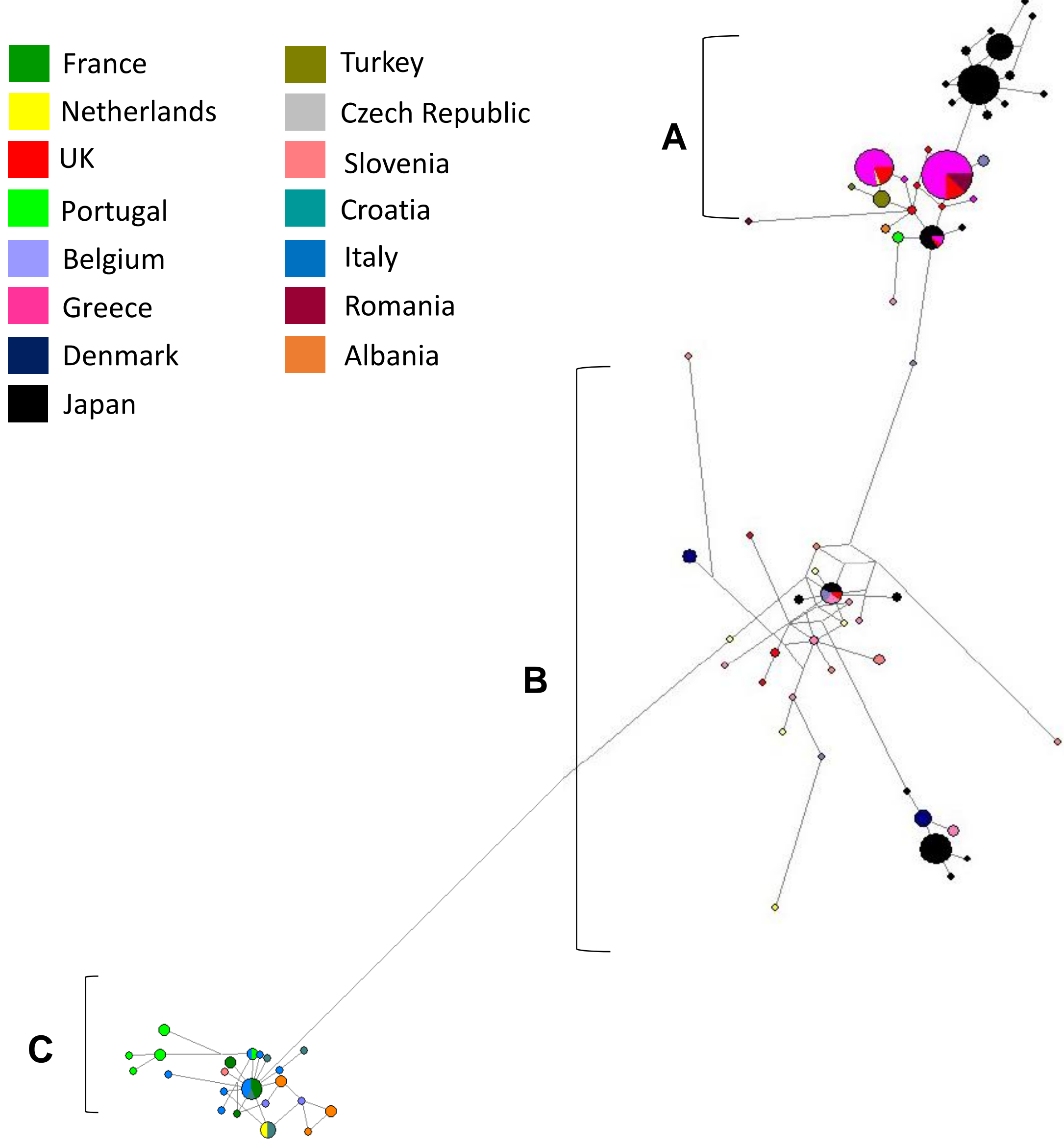


Figure 23: *Network analysis of all European and UK D. gallinae individuals sequenced in the study and Japanese sequences available from Genbank (237). Median-joining tree analysis was completed on Network 5.0.0.3. The three main haplogroups are labelled A, B and C. A total of 554 positions were used in this analysis, encoding 270 nucleotide sequences. Colour coded key provided for country identification.*

4.5 DISCUSSION

In this study, the phylogeny of *D. gallinae* populations was assessed by sequencing mitochondrial COI gene amplicons from 82 farms from 13 mainland European countries and the United Kingdom, including seven countries not previously studied in the published literature.

Previous research focusing on COI diversity in *D. gallinae* has demonstrated multiple lineages with comparative analyses concluding that cryptic species must be present (44, 234, 236). In the present study, multiple lineages were found with three main haplogroups (A, B, C). The C group haplotypes branched earlier in the phylogenetic tree when compared to groups A and B.

4.5.1 Geographical clustering

In some cases, as one might expect, phylogenetic clustering between countries sharing a border or located closely geographically can be seen. This was demonstrated in haplogroup C where sequences from Italy, Croatia, Albania, France, Slovenia, the Netherlands and Portugal clustered, providing support for hypothesis one that shared haplotypes would be observed between countries sharing a border. However, sequence analysis did reveal variation in nucleotide and haplotype diversity when looking at countries grouped by geographic distance (Figure 13-14). Analysis focusing on Belgium and the Netherlands (cluster two) showed no significant difference in nucleotide diversity compared to the full dataset, with only 0.00121 between Pi values, reflecting a high nucleotide diversity for cluster two (Table 12, Figure 13). Haplotype diversity was signifcantly higher for cluster two compared to the full dataset (Table 12, Figure 13). Conversely, groupings of Greece, Albania and Turkey (cluster five) and Portugal, France and Italy (cluster three) showed significantly lower nucleotide and haplotype diversity. In addition, samples spread across a greater geographical distance (e.g. Denmark to Slovenia) were clustered in haplogroup B. Network analysis illustrated the occurrence of shared haplotypes between multiple European countries (e.g. Belgium and the Czech Republic in Figure 21) and, in conjunction with comparative analysis between UK, mainland European and Japanese samples (Figures 22,-23), which supports previous evidence of international and intra-national movement of mites (237). In their study, Marangi et al. (2009) found higher nucleotide diversity in UK samples compared to Italian and French and suggested this highlighted levels of environmental pressures in each country (234). Across European countries, differences in control measures (and regulations restricting or permitting use of specific products), temperature differences, primary production systems and movement/trade of live hens and eggs might help explain variations in nucleotide and haplotype diversity. Greater knowledge in

these areas linked to data on genetic variation would help to form a better understanding of why differences are observed.

At present, it does not seem feasible to predict *D. gallinae* diversity based on gegraphical location and our phylogenetic analysis shows examples of geographical clustering, geographic diversity and non-geographical clustering. For future investigations the addition of data from connecting countries (i.e. Spain, Bosnia and Herzegovina and Montenegro) and higher farm sample numbers per country would help develop a clearer picture.

4.5.2 Genetic diversity of COI in UK *D. gallinae* populations

Establishment of *D. gallinae* populations from limited numbers of individuals is anticipated to have consequences on level of genetic diversity. Expansion from a small number of mites is likely to result in a relatively smaller number of haplotypes than expansion from a larger number of initial mites (235). In the UK, it was clear that despite being a group of islands the mite populations sampled were genetically related to those found in mainland Europe and Japan with nine haplotypes spread through haplogroups A and B (Figure 20) and three shared haplotypes between England and Japan (Figure 22). Identical sequences were found in one haplotype originating from the UK, Japan, Belgium and the Czech Republic, and in another haplotype from the UK, Japan and Greece. It seems most likely that trade between countries, either historical or on-going, provides opportunities for admixture of mites originating in different countries resulting in shared haplotypes. This provides phylogenetic evidence in support of hypothesis two, that shared haplotypes would be observed between the UK and mainland Europe.

Within the UK, phylogenetic analysis demonstrated that individuals from Scotland, England and Wales were distributed in both haplogroups A and B but only one haplotype was shared individuals from all three countries, found in haplogroup A (Figure 20). All other haplotypes contained individuals from a single country. Compared to the full dataset, nucleotide diversity was signficantly lower in the UK (Figure 9). Statistical analysis demonstrated the UK has significantly higher diversity than six countries, lower diversity than five countries and no signficant difference to Denmark and the Czech Republic). In contrast, the UK was not significantly lower in haplotype diversity than any country, or compared to the full dataset and significantly higher than five countries. Denmark was the only country to have no significant difference in nucleotide and haplotype diversity, suggesting a similarity in these parameters in both countries.

4.5.2.1 Differences in COI genetic diversity in UK production systems

Phylogenetic analysis of UK individuals demonstrated shared haplotypes between intensive and free-range layer farms (Figure 19). Individuals from intensive layer farms were present in two of the ten haplotypes in haplogroup A and three of the seven haplotypes in group B (Figure 20). Out of the 17 haplotypes from the UK, three contained a mixture of individuals from free-range and intensive layer farms, two in haplogroup one and one in haplogroup two (Figure 19).

Focusing on nucleotide and haplotype diversity, a small number of more diverse genotypes were found in free-range farms compared to intensive farms. Comparison of intensive caged systems with free-range systems revealed significantly higher nucleotide diversity for intensive systems but significantly higher haplotype diversity for free-range farms (Table 13, Figure 15) Two of the five farms displaying intra-farm variation were intensive systems and, of the seven nucleotide positions that showed variation from a single farm, three of these were attributed to a single intensive style farm (UK15). Despite a lower nucleotide diversity, overall free-range farms accounted for 60% of variable nucleotide positions in the UK with intra-farm variation (Table 17, Table 18, Figure 18).

Differences in nucleotide and haplotype diversity across production systems might be attributed to several factors. Outdoor access given to free-range laying hens provides a larger environment for parasite persistence and for transmission to occur (362) and, typically, higher *D. gallinae* mite populations are seen in free-range systems when compared to cage units (24, 26). In this respect, free-range systems are harder to clean in the same way as an intensive system. Higher mite populations could provide a plausible explanation for an increased haplotype diversity due to greater admixture occurring within farms and potentially increasing the frequency of recombination events. Other factors, such as number of barns/flocks kept at individual farms, whether workers cross between barns (potentially carrying *D. gallinae* across) and the nature of selection under each system may also help to explain the differences. The nature of selection at free-range and intensive farms can vary relating to differential use of control measures (e.g. acaricides) and this could play a role in shaping different levels of nucleotide and haplotype diversity.

It should be noted that low numbers of individuals were sequenced per farm and that an increase in the number of individuals sequenced would provide better clarification on the role production system plays on the genetic diversity of *D. gallinae*. Phylogenetic analysis and comparisons of haplotype and nucleotide diversity show that it is logical to reject hypothesis

three, that there will be no significant differences in COI diversity between UK production systems.

4.5.3 Intra-farm genetic variation

Intra-farm variation was observed in all four of the Greek farms sampled (Table 19), where three farms (Thessaloniki, Leros and Attica) had three haplotypes and one farm (Corinth) had two haplotypes. All of the haplotypes were assigned to haplogroups Aa and Ab and the two haplotypes from Corinth were shared by all three other farms (Figure 20). These two haplotypes represented the majority of individuals sampled from Greece, totalling 58 of 61. However, the third haplotype for Leros, Thessoliniki and Attica was individual to each farm, and, interestingly, shared an identical sequence with an individual originating in the UK (Figures 20-21).

Similar results were demonstrated for two farms investigated by others in Norway, where two and three different haplotypes were discovered from 17 and 19 individual *D. gallinae*, respectively (235). These authors reasoned that multiple haplotypes in a single farm is indicative of the farm either being infected by multiple haplotypes together or experiencing multiple infection events, stating that mite populations with contact have an increased chance of shared haplotypes than those with barriers separating them. In cases where haplotype occurrence could not be explained by geographical location they likely result from contaminated equipment, infected chickens or other materials being moved between farms. The scattering of haplotypes found in the present study is suggestive of the latter being true, that shared haplotypes could result from infected chickens or materials. Three of the farms sampled were located on the Greek mainland and the final farm was located on Leros, one of the islands in the Aegean sea. Despite being separated by the Aegean sea, all four farms shared two haplotypes, suggesting a common original source for all farms or continuous admixture between them. That would be possible by transport or trade routes, or sharing of contaminated equipment. This is also exemplified when considering that the common haplotype for all Greek farms found in haplogroup Aa also contained individuals from the UK.

Comparing UK farms, 60% of farms with multiple individuals sequenced showed intra-farm variation, suggesting that some farms have limited population diversity. However, it is worth noting that this could be related to low numbers of mites sampled per farm. A range in intra-farm variation can also be seen, with farms displaying variation at between 8-56% of the 27 variable base-pair positions (Table 17). Two farms showed intra-farm variation at 56% of the identified variable sites, UK14 and UK3, both free-range and located in England and Wales, respectively. Intra-farm variation was observed in English, Scottish and Welsh farms but not seen

in the farm from Northern Ireland. Further sequencing of mites from Northern Irish and Irish farms is required for an understanding of intra-farm variation present in the region. Similar to Greece, six UK farms demonstrated two to five haplotypes, with the remaining nine farms having one haplotype (however, five of these only had one individual sequenced so default to one). Of the UK farms, eight farms shared a haplotype with at least one other farm, with one haplotype in group one (Figure 19, Figure 21) consisting of individuals originating from five different farms. These five farms, from England, Scotland and Wales, were geographically spread aross the UK, further highlighting the possible spread of *D. gallinae* via transport links, infected poultry or other materials moved between farms.

4.5.4 Neutrality testing

Using neutrality theory, a population which adheres to a standard neutral model will have a Tajima's D value of zero (351). Overall, Tajima's D for 12 countries (including both production systems: intensive and free-range for the UK) and the whole dataset were not significantly different from 0, indicating that for these countries COI is neutral. These findings are in line with other research on the COI gene in *D. gallinae* conducted by Øines and Brännström (2011). They found a Tajima's D value of -1.25 for 46 *D. gallinae* individuals from Norway and Sweden and the result was not significant, suggesting a neutral population (235). In contrast, results from Fu and Li's D and F neutrality tests showed the full dataset to be significantly low for both D and F statistics ($P<0.02$) (Table 15), suggesting there is an excess of low frequency alleles, indicating population expansion or positive selection occuring (351, 352).

Greece and Denmark significantly differ from zero with positive values (2.83215 and 2.19756, respectively) for Tajima's D, indicating that these populations have an excess of intermediate frequency alleles, consistent with a recent population contraction or balancing selection (351, 352). Personal communication from the field have provided no indication of a population contraction, supporting the latter. Significantly positive values were also observed for the Fu and Li's D and F statistics for Denmark ($P<0.02$), but only for the F statistic for Greece ($P<0.05$) where the D statistic was positive but not significant ($P<0.10$) (Table 15). Portugal showed no signficiant difference for Tajima's D but had significantly positive values for Fu and Li's D and F statistics ($P<0.02$).

It is unclear why Greece, Denmark or Portugal would be experiencing balancing selection. One plausible explanation is the differential use of control measures across Europe placing different selection pressures on mite populations. However, the value of Tajima's D is known to be senstive to factors besides selection (363), including small sample sizes (364), recent bottlenecks

in the population which cause inflation of the number of θ relative to π (346) or a population subdivision, causing inflation of π relative to θ (350, 365). With only nine individuals for Denmark, it is plausible that low sample size has impacted on the Tajima's D score, however 61 individuals were included in the analysis for Greece.

4.6 Conclusions

This study provides evidence for genetic diversities in *D. gallinae* distributed across Europe. Where sufficient sequence depth was generated, intra-farm variation was detected in the United Kingdom and Greece. In addition, phylogenetic analysis provided further support for international and intranational movement of *D. gallinae*. Mapping additional COI diversity in countries not yet researched would help to build a more comprehensive understanding. Assessment of additonal nuclear genetic markers can be expected to provide further detail, and will be addressed later in this thesis.

5 GENOME-WIDE GENETIC ANALYSIS: SNP DISCOVERY THROUGH THE GATK PIPELINE

5.1 INTRODUCTION

5.1.1 Genetic diversity

Genetic diversity, also known as genetic polymorphism, can be defined as the variation in DNA sequence between individuals, or distinct chromosomes, of a given population or species (366, 367). Genetic diversity provides a crucial source of phenotypic diversity (368) and raw material for evolution via natural selection (369, 370). Genetic variation is, in fact, one of the three biodiversity levels recommended for conservation by the World Conservation Union (IUCN) (371). In nature there is abundant evidence for evolution by natural selection, confirming the role genetic variation plays for traits which influence fitness (e.g. (372, 373)) resulting in individual genotypes which vary ecologically. Genetic variation can arise as a result of expansion and/or contraction of short and variable tandem repeats, recombination events (such as chromosomal rearrangement), single-nucleotide polymorphisms and insertion or deletion polymorphisms (indels) (374-378). A lack of genetic diversity is typically associated with small or declining populations (that can be potentially endangered) (379).

In theory, genetic diversity in coding regions can be viewed as the balance between the appearance and disappearance of genetic variants, known as alleles (366). At each generation, new alleles will appear through spontaneous mutation resulting from errors in DNA replication or mutagen-induced damage (366). Mutation rate is not consistent across the genome (380) or among species (381) which can explain some of the observed variation in genetic diversity. The level of genetic diversity in a species is also ruled by the rate of allele fixation and loss, and loci with neutral alleles (i.e. those that have neither positive or negative effects and do not alter an individual's fitness) will largely be influenced by genetic drift (i.e. random fluctuation of frequency in alleles over generations) (382). Considering a panmictic population, also known as a Wright-Fisher population, where individuals are expected to equally contribute to reproduction, the strength of genetic drift will be inversely proportional to population size (366, 367). In reality, populations will depart from Wright-Fisher assumptions in a number of respects, thus the concept of effective population size (N_e). The N_e is the size of a population which would demonstrate the same level of genetic diversity as the population of interest (366). As a result, the neutral theory of molecular evolution (383) predicts that a population of constant size will have a genetic diversity that is proportional to N_e (366).

Much of the genome is non-coding and as such most DNA sequence variation occurs in non-coding regions (384). Non-coding elements include *cis*-regulatory elements (silencers, enhancers, promoters and insulators) as well as non-coding RNAs (ncRNAs) (385). Non-coding

RNAs have been identified as critical regulators of multiple pathways and cellular mechanisms (386-388), as well as being implicated in several diseases (389, 390). As in coding sequence (CDS), non-coding regions experience disruption through single nucleotide polymorphisms, larger structural variants and indels (385).

5.1.1.1 The neutral theory of molecular evolution

Under the neutral theory, a population's genetic diversity is dependent on the N_e and the mutation rate (383). The theory states that genetic diversity levels at neutral sites are a reflection of the balance between mutational input and a loss in genetic diversity due to random sampling in a finite population ("*genetic drift*") (383, 391, 392). At present, most evolutionary research uses neutral expectations for their default null model (393, 394). Under the assumption that everything else remains equal, following neutral theory, the level of genetic diversity within a species is anticipated to proportionally increase in line with the census population size, N_c, or the number of breeding individuals (395). A large body of theoretical and empirical work has shown that neutral genetic diversity (i.e. that which is not affected by natural selection) within a species can be influenced via nearby genetic variants which are affected by natural selection (which is reviewed in (26)). Within a population, the level of neutral genetic diversity is central in aiding the understanding of conservation genetics (396), genome-wide associations with disease (397), demographic histories of populations (398, 399), selective constraints (400), and the molecular basis of adaptive evolution (401).

As time has passed, research has strongly challenged some of the key assumptions from the neutral theory of molecular evolution (402). Firstly, in a number of species adaptation rates appear to be high, for example in *Drosophila melanogaster* more than 50% of amino-acid changing substitutions and an approximately equal proportion of non-coding substitutions are driven to fixation through positive selection (403). The apparent lack of a clear relationship between population size and genetic diversity observed in wild populations has led some authors to argue that genetic drift and population size are not major factors that impact molecular variation (404-408). Importantly, it is suggested that adaptation commonly affects patterns of genome-wide polymorphisms (403, 409-411), implying that a given polymorphism's dynamics are not just affected by genetic drift and purifying selection but also influenced by genetic draft (412). Genetic draft, also known as genetic hitchhiking or the hitchhiking effect, is used to describe the stochastic effects at closely linked sites generated through selective sweeps (402), that is, the changes in allele frequency as a result of partially heritable random associations with genetic backgrounds (413). Secondly, there is an accumulation of evidence showing that many polymorphisms occurring in natural populations are mildly deleterious (414-

418), and as such, these polymorphisms are anticipated to generate another form of interference among linked sites, commonly referred to as background selection (418, 419). Despite the on-going dispute over the neutral theory (383) and its validity, it has still been one of the dominant frameworks used in population genetics and evolutionary research for over 40 years (399, 420).

5.1.1.2 Determinants of genetic diversity

A wide range of factors influence genetic diversity, but overall, effective population size (N_e), mutation and linked selection are the main factors. These factors are all ruled by a number of other parameters, including population demography (migration or bottleneck events), mating system (including asexuality), recombination, gene density and a population's life history (including lifespan and fecundity) (366). The connections between these parameters and their correlation to genetic diversity (both positive and negative) can be seen in Figure 24. The majority of genetic variants associated with complex traits lie in non-coding regions of the genome, with many lying some distance away from the nearest protein-coding locus (421). This observation implies that many variants affecting the risk of common, complex diseases are likely to exert their effect by altering the regulation of genes rather than by directly affecting gene and protein function (422).

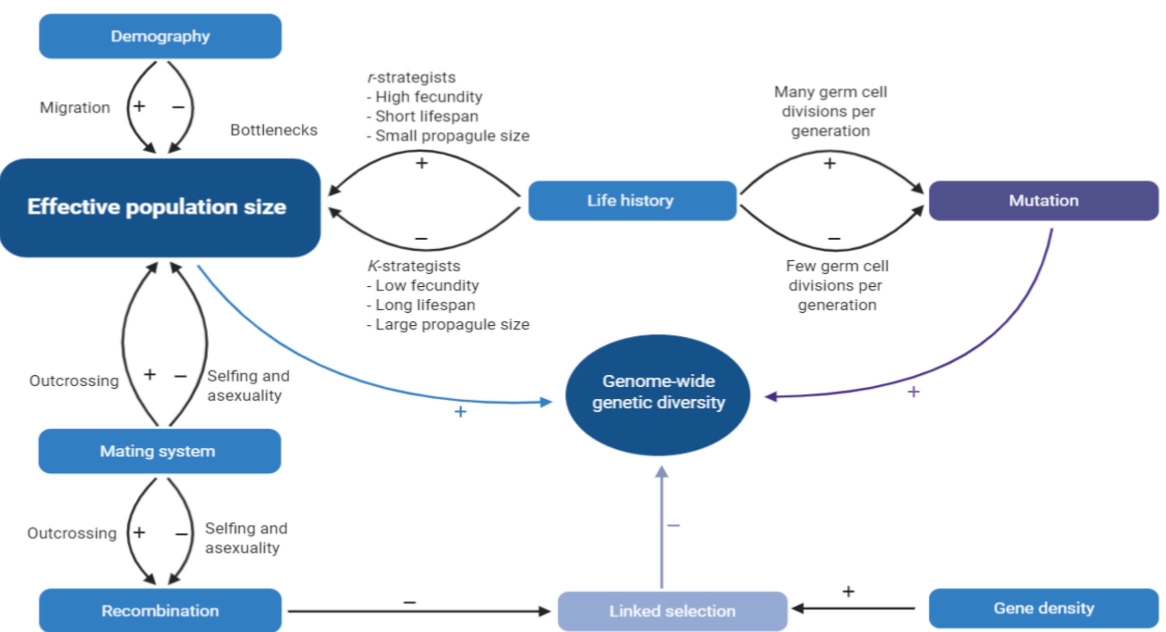


Figure 24: *Overview of determinants of genetic diversity. Effective population size, mutation rate and linked selection are the main factors affecting diversity. These factors are governed by several other parameters. The direction of the correlation is indicated by the + and – symbols. Selfing – self-fertilization. Adapted from (366)*

5.1.2 Quantifying genetic diversity

Genetic diversity is typically characterised by using data depicting variation in either discrete allelic states (e.g. allelic diversity, allelic or genotypic richness, heterozygosity, nucleotide diversity etc) or continuously distributed (i.e. quantitative) characters (e.g. coefficient of genetic variance, genetic variance and heritability), which lead to different possible genetic diversity metrics. This variation in phenotypic traits or allelic traits can be neutral or non-neutral in regards to consequences on fitness (368). For example, molecular markers, including amplified fragment length polymorphisms (AFLPs), microsatellites, protein polymorphisms or direct DNA sequences (305) normally represent discrete allelic states which are assumed neutral. Whilst typically neutral traits are measured as discrete allelic states, not all of these traits measured are neutral (372). Theoretically, quantitative traits can also be neutral but in the majority of cases researchers have focused on traits with a presumed or known functional significance (368). Focusing on discrete allelic states, discrete metrics of genetic diversity represent either haplotype frequency or the number of alleles (423). Haplotype frequencies can be measured through estimation of the probability that two randomly chosen haplotypes or alleles in a population can be expected to differ, given certain assumptions about the ecology and genetics of the populations (e.g. gene diversity, expected heterozygosity etc.) (368). In general, discrete trait metrics used to depict genetic diversity are tailored in accordance with the type of inheritance for a particular genetic marker (424) however, overall, they typically reflect either the richness (number) and/or evenness (relative frequency) of alleles .

5.1.3 Single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) are considered an abundant form of genome variation which can be defined as a single base pair position change at a specific location in the genome (425-427). Typically, they are distinguished from rare variants by a requirement that the least abundant allele must have a frequency of 1% or greater (425, 428, 429). Thus, under formal considerations, a single base insertion or deletion (indel) would not be considered a SNP. However, in practice a strict definition is not always applied and in some cases biallelic variations including deletions, insertions and variants with a allele frequency under 1% are referred to as SNPs (425). In theory, SNPs can be bi-, tri, or tetra-allelic variations (428), however tri- and tetra-allelic SNPs are uncommon and typically the majority of SNPs are biallelic (425, 430). Transitions, such as A ⇔ G or T ⇔ C, have a higher prevalence than transversions, such as A ⇔ T or C; and G ⇔ T or C (374). There are twice as many possible transversions compared to transitions, such that the transitions over transversion ratio should be 0.5 if mutations occur randomly. Despite this, observed data demonstrates a clear bias towards transitions (377). Research indicates that

the majority of SNPs are found in non-coding regions of the genome (431). Although SNPs in noncoding regions do not alter encoded proteins, they can influence gene regulation, and they serve as important physical or genetic markers for both evolutionary and comparative genomics research (425).

Various types of SNPs can influence the function or regulation and expression of DNA, RNA and proteins, and their classification is typically based on genomic location (426, 432). Non-synonymous SNPs are those in coding sequences where the amino acid sequence of the protein product is altered (433), either through introduction of a non-sense or truncation mutation, or through amino acid substitution (374). SNPs can affect the translation or expression of gene products through interruption of regulatory regions or through interference with normal splicing and mRNA function, including synonymous SNPs (i.e. one that does not alter amino acid sequence), and SNPs located in regulatory regions or intronic SNPs (426). SNPs that are polymorphic at splice sites can result in variant proteins which differ in the exons that they contain (432). In promoter regions, SNPs have been reported to affect the expression and regulation of proteins which directly impact protein function (434). Mutations that result in premature stop codons can lead to production of truncated protein products or near-null phenotypes due to nonsense mediated decay (374, 435, 436). Several reviews are accessible for understanding the mechanisms by which SNPs affect protein structure and the use of SNPs for detection of functional variants in candidate genes and use in genetics studies (377, 437).

SNPs have been widely employed as molecular markers in genetic improvement of plants and animals. These applications include high-resolution genetic map construction, linkage disequilibrium based association mapping, genetic diagnostics, genetic diversity analysis, cultivar identification, phylogenetic analysis and characterisation of genetic resources (432, 438). The usefulness of SNPs in analyses of population diversity and structure has been demonstrated in several studies (439, 440) and analysis of genome-wide markers in both populations and pedigrees has been essential for the evaluation of processes and patterns underlying evolutionary change and for investigation into the genetic architecture that underpins quantitative and other phenotypic traits (441).

Once discovered, SNPs can be converted into genetic markers that can be inexpensively assayed in a high-throughput manner (442, 443), and it is possible to use SNP-based markers to generate dense genetic maps (444). Such maps can be used to conduct marker-assisted selection (MAS) programs, construct the specific genotypes required for quantitative genetic studies, to enhance our understanding of genome organization and function and address fundamental questions relating to evolution, segregation and meiotic recombination (445). Furthermore transcript-

associated SNPs can be used to develop allele-specific assays for the examination of cis-regulatory variation within a species (446-449).

5.1.3.1 SNP discovery process

SNP discovery is the process of identifying polymorphic sites in the genome of a species and/or population of interest (427). Multiple methods for SNP discovery have been proposed (294, 450-453), but typically they rely on scoring schemes which identify SNPs once the score exceeds a set threshold (454). In humans, the majority of SNP discovery has been done *in silico*, such that individuals in public databases were screened to allow for the identification of putative polymorphisms (e.g. 9) (427). For non-model organisms, SNPs are typically identified via laboratory screening (e.g. sequencing) or through segments of the genome taken from multiple individuals. The total number of genome segments required for discovery of a predetermined number of SNPs is dependent on the SNP density across the genome (427). For many species, SNPs are found to occur every 200-500bp, implying that screening 75-100 genome segments of approximately 500-800 bp should result in identification of more than 50 independent SNPs (4,12). One cost efficient and fast approach for identifying large numbers of SNPs is through data mining large scale sequencing projects and using public repositories of sequence data (455). Multiple sequence alignments or pairwise alignments can be utilised in the evaluation of individual bases and SNP identification can result by comparison of sequence variants that are represented by multiple reads (432). The greater the number of available sequence reads which represent a specific genomic location, the higher the chance of identifying robust polymorphisms (456). It is possible to distinguish a sequencing error from a sequence variant when it is confirmed by multiple reads and typically there is a greater probability of it being a true polymorphism when a higher number of reads per allele are available (432, 456). Successful identification of SNPs through comparison of genomes from two or more genetically distinct individuals is evidenced in the literature, from inbred lines of mice (457), the *japonica* and *indica* *s*ub-species of rice (458), different lines of maize (459), and the Landsberg and Columbia ecotypes of *Arabidopsis* (460).

5.1.3.2 Utilising bioinformatics for SNP identification

From basic research through to translational genomics in clinics, the analysis of variant data as a result of genome or exome sequencing has been fundamental for progression in biological understanding (461). The development of high-throughput sequencing platforms means that it is now possible for individual laboratories to generate DNA sequence at unprecedented rates and consequently allow for the compilation of huge amounts of genetic data (219). However,

this massive amount of data constitutes a substantial issue for downstream studies because the analysis and interpretation rely on highly specialised software and expertise (219).

A range of bioinformatic tools have been developed over the past decade or so that are available for use in handling DNA sequencing data (219). These include tools for use in alignment of raw sequencing reads to reference genomes (221), such as BWA (462) and Bowtie (223), tools for assembling a new genome (224), such as FERMI (225), Abyss (226) or SoapDenovo (218), and tools which can perform data quality control (220). Alongside these, tools are available for calling single nucleotide variants (227), such as the Genome Analysis Toolkit (GATK) (228), and tools used in the calling of structural variants (229), including ERDS (231) and CNVnator (230).

Identification of a SNP is achieved when a nucleotide from an accession read is found to differ from a reference genome at the same nucleotide position (454, 463). When a reference genome is absent, SNP mining can be achieved by comparison of reads from differing genotypes through *de novo* assembly strategies (464). In principle, available methods generally encompass the following steps depicted in Figure 25 (456).

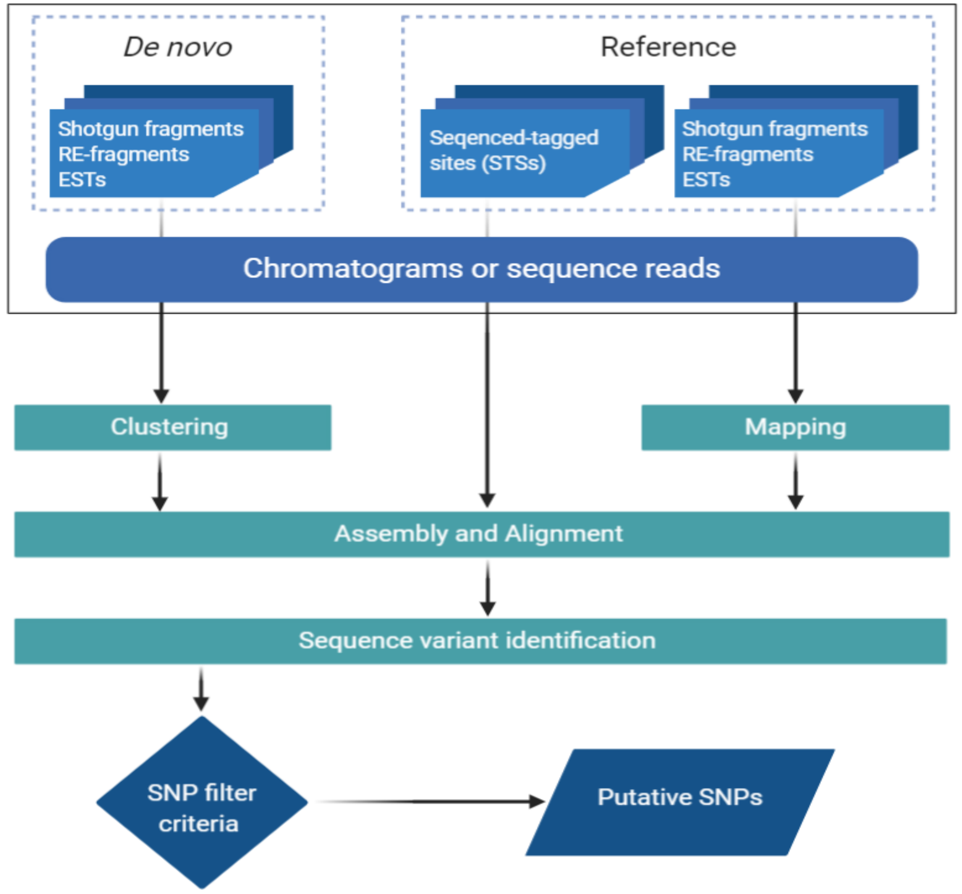


Figure 25: *Flowchart of the general approach to single nucleotide polymorphism (SNP) mining from DNA sequence data. STSs: sequence-tagged sites, ESTs: Expressed sequence tags, RE-fragments: Restriction fragments. Reproduced from (456).*

Read assembly files generated by mapping programmes are used to perform SNP calling (463). In practice, various empirical and statistical criteria are used to call SNPs, such as minimum and maximum number of reads considering the read depth, the quality score and the consensus base ratio for examples (464). Thresholds for these criteria are adjusted based on the read length and the genome coverage achieved by NGS data (463). In assemblies generated allowing single nucleotide variants and insertions indels, a list of SNP and indel coordinates is generated and the read mapping results can be visualised using graphical user interface programs such as Tablet (465), SNP-VISTA (466) or Savant (467).

5.2 AIMS AND HYPOTHESES

5.2.1 Main aim

The main aim of this study was to utilise bioinformatic tools and pipelines to identify candidate SNPs for downstream genotyping. Originally the aim was to achieve this through comparison of *D. gallinae* transcriptomic data sets but was adjusted to incorporate comparison of the *D. gallinae* transcriptome with the draft *D. gallinae* genome when it became available in late 2018.

5.2.2 Hypothesis

1. *D. gallinae* transcriptomic resources can be utilised in the identification of SNPs that can aid in defining the draft *D. gallinae* genome assembly

5.3 METHODOLOGY

5.3.1 Current resources for *Dermanyssus gallinae*

5.3.1.1 Transcriptomes

In 2014, Schicht *et al.,* published a transcriptome analysis of *D. gallinae* (468). Total RNA was extracted from an acaricide-susceptible strain of *D. gallinae* maintained at the University of Veterinary Medicine Hannover, Institute for Parasitology. Synthesis of cDNA was completed from a pool of both sexes and all developmental stages of starved mites and fed mites, with Roche 454 sequencing used to produce the transcriptome. The final dataset consists of 267,464 sequences (231,657 singletons, 56 contigs and 35,751 isotigs) (468).

The RVC have produced a transcriptome dataset utilising Illumnia NGS sequencing and transcriptomic data has been shared by the Moredun. The Moredun transcriptome project was

completed using 454 Roche sequencing from total RNA isolated from fed mites of mixed life stages and ages collected from layer-farms in Scotland. A total of 13,363 contigs and 325,432 singletons were produced, with 13,330 contigs retained after filtering. The RVC transcriptome project was produced from mixed stage, starved mites collected from a layer farm in England and one Illumnia flow cell lane which generated paired end 100bp reads. A total of 36,199 Mb of sequence was produced from ~361 million reads and assembled into >200k clusters. These transcriptomic data sets are not publicly available but were utilised in the generation of the draft genome assembly and during SNP validation.

5.3.1.2 Genome assembly

In 2018, Burgess *et al.,* released a draft genome assembly of the *D. gallinae* genome (232). They extracted genomic DNA from adult female *D. gallinae* mites and freshly laid eggs, collected from a layer farm in Scotland (same as Moredun transcriptome). They used a combination of PacBio and Oxford Nanopore Technologies MinION to produce a final assembly of 7,171 contigs and an assembled genome size of 959 Mb (232).

5.3.2 The Genome Analysis Toolkit

The genome analysis toolkit (GATK) is a structured JAVA programming framework designed to enable the development of robust and efficient analysis of next generation DNA sequences following MapReduce functional programming philosophy (228). All GATK analyses were completed by uploading sequence data to the Galaxy web platform and the usegalaxy.org public server was used for all analytical steps (469, 470). The GATK germline short variant discovery (SNPs and Indels) best practices pipeline (471) was followed as closely as possible, with a few adaptations as discussed later. The main steps of the pipeline can be seen in Figure 26 below.

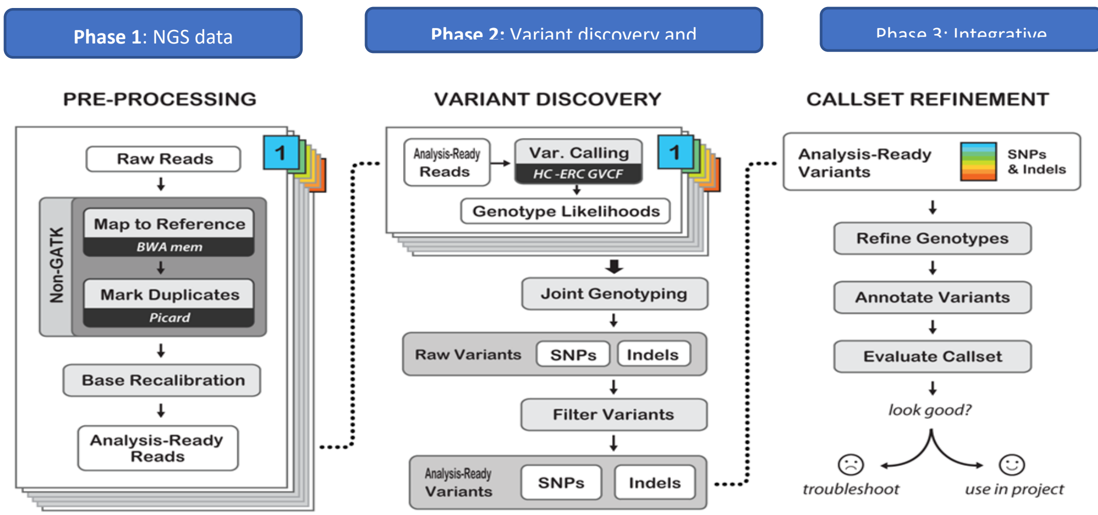


Figure 26*: GATK best practices for germline short variant discovery (SNP and Indels). Sourced from (471).*

The pipeline can be thought of as a three-part conceptual framework, as outlined by DePristo *et al.,* (472). Phase one involves the processing of next-generation data, phase two involves variant discovery and genotyping and phase three involves integrative analysis (472). For this pipeline, phases one and two were implemented.

5.3.3 Phase one

Phase one of the pipeline involved transforming platform-dependant biases into single, generic representations that had well-calibrated base error estimates. These were aligned to the correct genomic origin (i.e. the draft *D. gallinae* genome assembly) and aligned consistently with respect to each other (472). Mapping algorithms use an initial alignment to place reads on the reference genome then either through generation or conversion produce a technology-independent BAM reference file format (473). Once completed, any molecular duplicates were eliminated and refinement of initial alignments was achieved through local realignment, after which empirically accurate per base error models were determined (472).

5.3.3.1 Raw reads: Input data

The transcriptome produced by Schicht et al. ("Hannover transcriptome") was uploaded onto the usegalaxy.org platform as two read sets in raw read format (Table 20)(469). The draft genome assembly was also uploaded to the usegalaxy.org platform and was set as the 'reference' genome.

Dataset	Read sets	File size	Sequencing platform	Read type	Quality scores
Hannover transcriptome (468)	1	12.9MB	454	Single	Y
	2	13.5MB	454	Single	Y
Draft genome assembly (232)	1	959MB	PacBio and Minion	Single	Y

Table 20*: Input genomic and transcriptomic* D. gallinae *data utilised in the SNP discovery pipeline. Number of read sets, file size, read type and sequencing platform, and whether quality scores were available for each dataset, are indicated.*

The Hannover transcriptome read sets were treated independently for SNP discovery. The two read sets were generated from two distinct 454 sequencing runs from the same biological material (468), and as such were treated as separate data sets. Read sets one and two were run through the GATK pipeline in separate runs, rather than simultaneously in one flow-through, for identification of SNPs before merging of VCF files.

The RVC and Moredun transcriptomes were planned to be used in the same pipeline, but due to time constraints were not included.

5.3.3.2 Non-GATK steps

5.3.3.2.1 NGS: QC and Manipulation tool

Raw reads from the transcriptome datasets had to be converted to a modified FASTQ format for GATK. To achieve this, the transcriptome FASTQ files were processed through the NGS: FASTQ Groomer tool available on galaxy with no advanced options selected (found under genomic file manipulation -> FASTA/FASTQ).

5.3.3.2.2 Mapping with BWA-MEM

Individual read sets from the transcriptome were selected to be mapped to the genome assembly with BWA-MEM (Burrows-Wheeler Aligner Maximal Exact Matcher). BWA-MEM is an alignment algorithm for alignment of sequence reads against large reference genomes (474). The algorithm for constructing the Burrows-Wheeler transform (BWT) index was set to Auto, allowing the BWA tool to decide the best algorithm to use. Mean, standard deviation, maximum and minimum insert lengths were not provided. Read groups were set to SAM/BAM specification. Read group (ID), read group sample name (SM) and library name (LB) were set to auto assign. Platform/technology used was set to the LS454. Sequencing centre that produced the read (CN), description (DS), data that run was produced (DT), flow order (FO), the array of nucleotide bases that correspond to the key sequence of each read (KS), programmes used for processing the read group (PG), predicted median insert size (PI) and platform unit (PU) were left blank. Simple Illumina mode was chosen for analysis mode. For the Job Resource Parameters, the default job parameters were selected.
BWA-MEM mapping results were analysed using Samtools flagstat tool on usegalaxy (475).

5.3.3.2.3 NGS Picard: MarkDuplicates

MarkDuplicates is a Picard tool which has the purpose of examining the alignment BAM dataset to identify duplicate molecules (476). BWA-MEM mapped files were selected, with no comment inserted. For the option "*If true do not write duplicates to the output file instead of writing them with the appropriate flags set*", no was selected. Assume the input file is already sorted set to yes. Sum_of_base_qualities was set for the scoring strategy for choosing the non-duplicate among candidates. The regular expression that can be used to parse read names in the incoming SAM/BAM dataset set to: [a-Za-Z0-9]+:[0-9]:([0-9]+):([0-9]+):([0-9]+).*. The maximum offset between two duplicate clusters in order to consider them optical duplicates set to 100. Lenient validation stringency allocated.

5.3.3.3 GATK steps

5.3.3.3.1 Base Quality Score Recalibration (BSQR) procedure

Base recalibration consists of three steps of data processing 1) counting and analysis of covariates on BAM files 2) table recalibration and 3) recounting of analysis of covariates on recalibrated BAM files.

5.3.3.3.1.1 Count covariates

Count covariates on BAM files is a walker acting as the first pass in a two-pass processing step of the GATK pipeline. The walker operates through a by-locus transversal that operates at sites not found in a dbSNP ROD file (single nucleotide polymorphism database reference-ordered data file) or a VCF file containing known variants (477). The walker works on the assumption that any reference mismatches observed are errors and indicate poor base quality. Empirical quality is calculated by p(error) = number of mismatches/number of observations, outputting a table of several covariate values (477).

5.3.3.3.1.1.1 Complications with D. gallinae dataset

GATK states that "*This calculation is critically dependent on being able to skip over known variant sites. Please provide a dbSNP ROD or a VCF file containing known sites of genetic variation." However, if you do not provide this file, the '--run_without_dbsnp_potentially_ruining_quality' flag will be automatically used, and the command will be allowed to run*". The GATK pipeline process was started in 2017 and completed in 2018. At this time, there was no database of SNPs in *D. gallinae* or VCF file containing known variants available to be used, presenting a complication in pipeline progress.

5.3.3.3.1.1.2 Solution to lack of known variants

Based on recommendations from the GATK website and forums, a self-validating GATK pipeline was curated, involving four steps (Figure 27).

1) Initial round of SNP discovery calling with no VCF file provided for recalibration forcing the *--run_without_dbsnp_potentially_ruining_quality* flag and no variant recalibration step completed
2) Intersection of VCF files produced from both read sets and removal of any SNPs that had a quality score lower than 20
3) Second round of SNP discovery, providing the VCF file curated in step two for base recalibration

4) Production of the SNP table

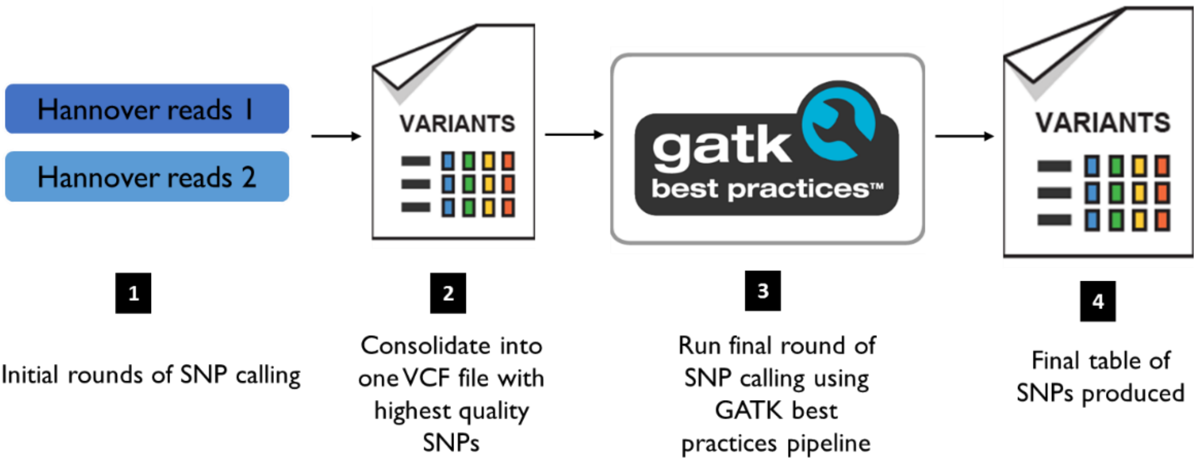


Figure 27: *An overview of the process used for SNP identification in D. gallinae using GATK. The four key stages involved are outlined 1) Initial round of SNP identification 2) Consolidation of VCF files 3) A second round of SNP calling and 4) Production of final SNP tables*

5.3.3.3.1.1.3 Count Covariate parameters

BAM files post MarkDuplicate process were selected as input BAM file, with the genome assembly used as the reference. "*Use the standard set of covariates in addition to the ones selected*" option was set to yes. From the list of '*Covariates to be used in the recalibration*' the following four were selected: ReadGroupCovariate, QualityScoreCovariate, CycleCovariate and DinucCovariate. No binding for reference-ordered data file attached (for reasons previously discussed). Basic GATK and analysis options were selected and default job resource parameters were used. Basic GATK options as stated here and in the following methodology sections refers to terminology used on the usegalaxy.com platform options. Basic or advanced options are available for use, with basic options essentially representing default parameters for the GATK pipeline and advanced options allowing the addition of user provided files or parameter changes. In cases where advanced options were not necessary, basic options were selected.

5.3.3.3.1.2 Analyse covariates

The analyse covariates tool generates plots which are used to assess the quality of a recalibration run, providing part of the Base Quality Score Recalibration (BSQR) procedure (478). Count covariate files were selected from the previous step, selected with basic options run.

5.3.3.3.1.3 Table recalibration

The table recalibration walker acted as the second pass in the two-pass processing step, achieved through by-read transversal. The walker calculated various covariates for each base in each read and then used these values as a key in a hashmap. The walker generated an empirical

base quality score, overwriting the quality score that was currently in the read and generating a new recalibrated BAM file. Count covariates files previously generated were selected for the recalibration file and the BAM files post MarkDuplicates process were selected for the input BAM file, with the *D. gallinae* genome assembly selected for the reference file. Default GATK and analysis options were selected.

5.3.3.3.1.4 Recounting and analysis of Covariates

In order to assess quality of recalibration, the Count Covariate and Analyse Covariate tools were re-run, except for using the recalibrated BAM files. For count covariates the same parameters outlined in 2.7.2.2.1.1.3 were followed and for analyse covariates basic options were selected, as done previously.

5.3.4 Phase two

In phase two, analysis reading BAM files are analysed for identification of all sites that had statistical evidence that an alternate allele was present, among samples including CNVs, short indels and SNPs (26). In this pipeline, the focus was purely on SNPs.

5.3.4.1 Variant Discovery

5.3.4.1.1 Unified Genotyper SNP and indel caller

The unified genotyper is a tool designed to unify the approaches of multiple disparate callers. For the BAM file, table recalibrated BAM files were selected and the genome assembly was used as the reference file. No binding for reference-ordered data file added. BOTH was selected as the genotype likelihood model to employ and a minimum phred-scaled confidence threshold at which variants not at ‘trigger’ track sites was adjusted to 20.0. The minimum phred-scaled confidence threshold at which variants not at ‘trigger’ track sites should be emitted (and filtered if less than the calling threshold) was also set to 20.0. For GATK options, the basic option was selected. For the analysis options, advanced options were chosen and everything was set to default parameters with the exception of annotation types and annotation interfaces/groups. For annotation interfaces/groups standard was selected. The following annotation types were selected: FisherStrand, HaplotypeScore, HomopolymerRun, MappingQualityRankSumTest, QualbyDepth and ReadPosRankSumTest (Table 21). No annotations were selected to be excluded.

Name	Summary
FisherStrand	Strand bias estimated through using Fishers exact test (FS) (479)
MappingQualityRankSumTest	Rank sum testing for mapping qualities of REF vs Alt reads (480)
QualbyDepth	The variant confidence after normalisation by the unfiltered depth of variant samples (481)
ReadPosRankSumTest	Rank sum testing for the relative positioning of REF vs ALT alleles within reads (482)

Table 21: *Summary of annotations applied in the UnifiedGenotyper step of the GATK pipeline for SNP identification in* D. gallinae*. The summaries of the following variants are outlined: FIsherStrand, MappingQualityRankSumTest, QualbyDepth and ReadPosRankSumTest.*

5.3.4.1.2 NGS: GATK – Select Variants from VCF files

The select variant tool was used to pull out SNPs from other variant types. Unified genotyper files were selected for the variant file and the genome assembly used for the reference file. No criteria to use when selecting the data applied, no output variants added, no samples included or excluded, and no filtered loci in the analysis. Basic GATK options were selected. Advanced analysis options selected, all parameters kept the same, except for selecting SNP as the option for "*select only a certain type of variants from the input fil*e".

This marked the end of the first round of SNP discovery.

5.3.5 VCF Intersection

To allow for identification of SNPs common to both read sets, intersection of the VCF output was undertaken. To complete this, the first VCF database input file selected was Hannover read set one and the second VCF database input file selected Hannover read set two. The genome assembly was used as the reference file. Intersection was selected, instead of union. Invert selection was not selected and records compared up to this many bp away (window size) set at 30. No was selected for output whole loci when one alternate allele matches, and advanced options were not utilised.

5.3.6 Second round of SNP calling

Pre-processing and variant discovery steps were completed as per the first round of SNP calling, except for the following changes:

1. Binding-reference ordered data file provided for all GATK steps that required it, using the intersected VCF file of Hannover 1 and Hannover 2 results from initial round of SNP calling (see 2.6.4).
2. Variant filtration was completed as described in the following sections.

5.3.6.1 Variant filtration

Variant quality score recalibration (VSQR) is a sophisticated filtering technique, applied to variant call sets, that utilises machine learning to be able to model variants technical profiles in a training set and uses this in order to filter probable artefacts from the call set (483). As part of the VSQR process, the variant recalibrator fitted a Gaussian mixture model to the contextual annotations provided for each variant.

5.3.6.1.1 Variant recalibrator

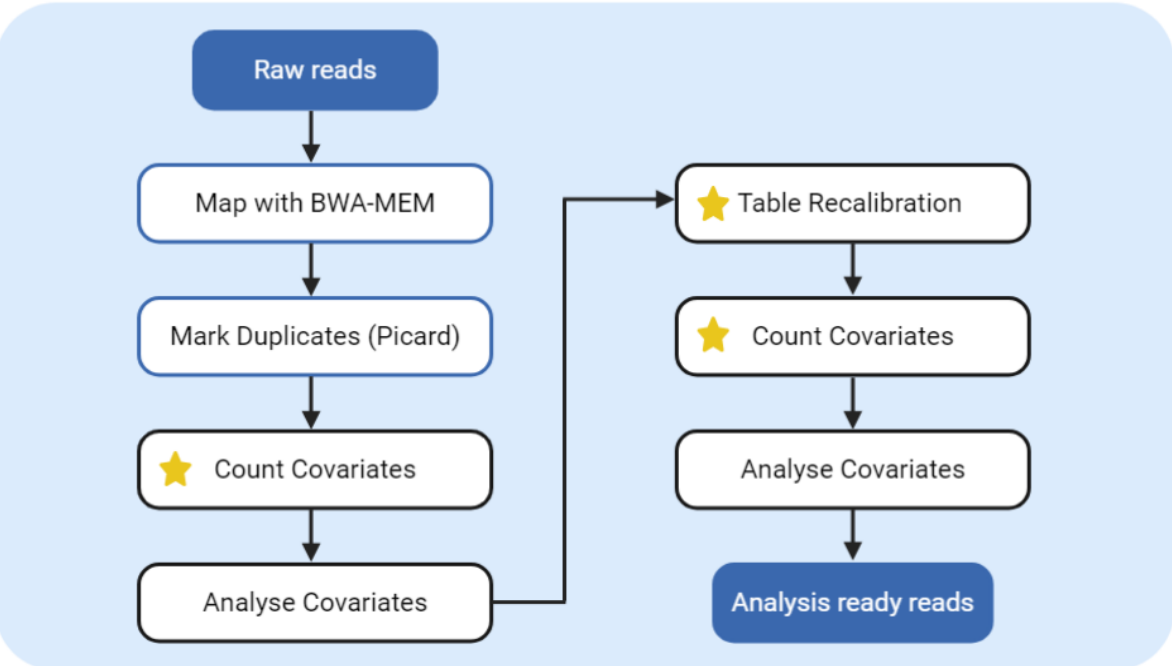
The variant recalibrator tool used the overlap between training/truth sets from the call set, modelled the distribution of variants in relation to the annotations selected, and attempted to form them into clusters. After clustering, it assigned a VQSOD score to all variants and variants that were at the heart of the cluster scored higher than variants which were outliers (483). For the variant file to recalibrate, the select variants variant file was selected with the genome assembly used as the reference and round 1 VCF output selected for the binding -reference ordered data option. All annotations for the "*annotations which should be used for calculations*" were selected, with a recalibration mode of SNP chosen. No additional annotations were included. Basic GATK options and basic analysis options were chosen, with default job resource parameters selected.

5.3.6.1.2 Apply Recalibration

The apply recalibration tool applied a filtering threshold to provide an indication of which variants passed filtration and which failed (483). The genome assembly was used as the reference file. For the "*variant file to annotate*", "*variant recalibration file*" and "*variant tranches file*" the appropriate output files were selected from variant recalibration. The genome assembly was used as the reference genome and basic GATK options were selected. The recalibration mode was left at SNP and no ignore filters were inserted. The truth sensitivity level at which to start filtering was left set at 99.0.

5.3.7 Summary of pipeline process

In summary, the GATK best practices for germline SNP and indel discovery was run a total of four times, twice for each transcriptomic read set. The entire process can be seen detailed in Figure 28, visualising the main differences between the first and second round.



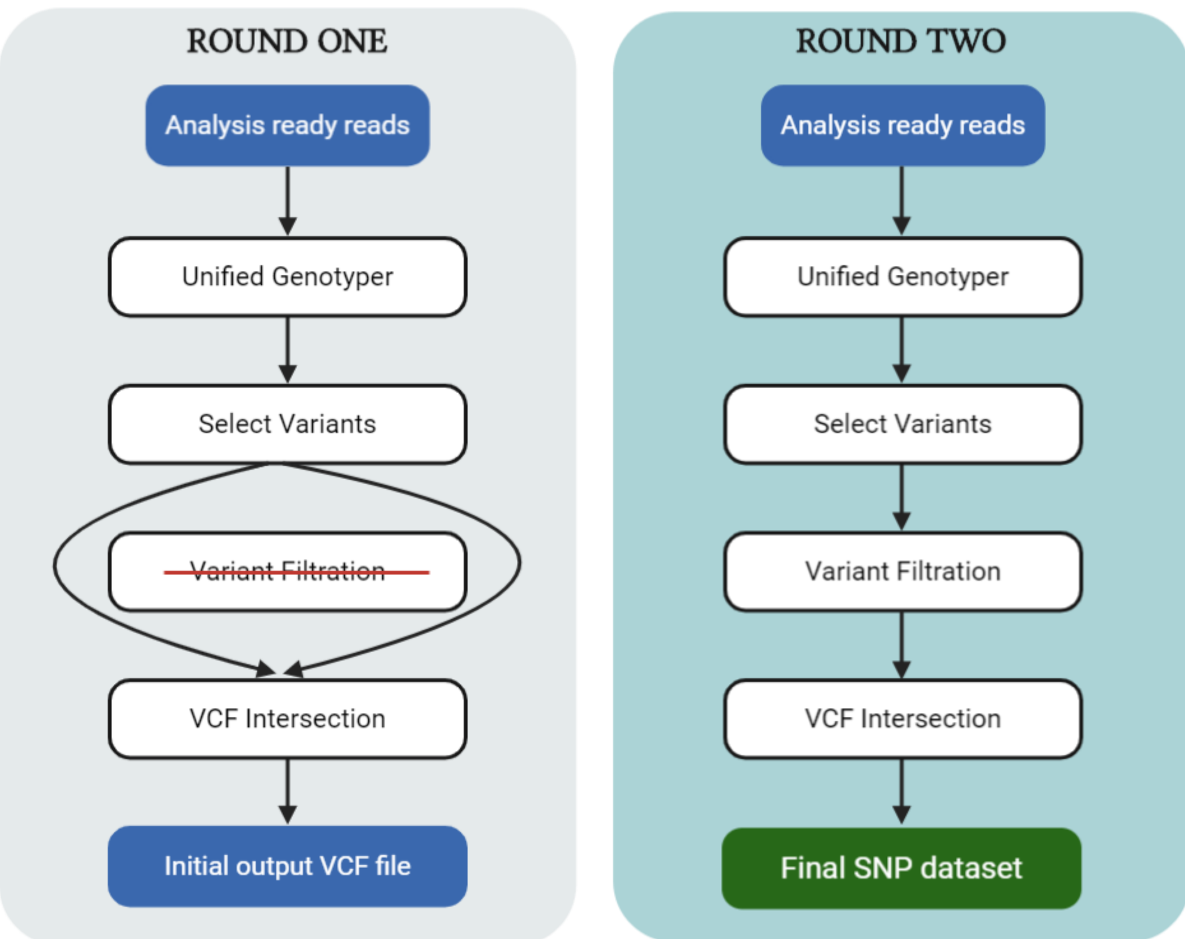


Figure 28: *Overview of the adapted GATK pipeline used for self-validation for SNP identification in* D. gallinae *Round one and two followed the same phase one, with the exception of yellow stars (*) indicating where reference SNPS were provided in round two. Phase two indicating differences in round one (grey) and round two (teal).*

5.4 Results

5.4.1 BWA-MEM mapping

BWA-mapping results using Samtools flagstat can be seen in Table 22, with a consistent result of 96.43% and 96.42% for read sets one and two over both rounds of SNP discovery.

Dataset	Read sets	File size	Mapping results Round 1	Mapping results Round 2
Hannover transcriptome	1	12.9MB	96.43%	96.43%
	2	13.5MB	96.42%	96.42%

***Table 22**: BWA-MEM mapping results for each transcriptomic read set to the D. gallinae genome assembly with percentage of raw reads according to 5.3.3.1 successfully mapped outlined for each round of SNP discovery*

5.4.2 SNP discovery round 1

A total of 130,532 SNPs were called from both transcriptome read sets when compared to the genome, 63,592 for read set one and 69,440 for read set two, with 96.43% and 96.42% of both read sets mapped to the genome (Table 23). For read set one, phred quality scores ranged from 30.02 to 3410.32, with an average score of 111.44. A total of 2,464 contigs were identified to have a SNP present, representing 34% of the genome. For read set two, phred quality scores ranged from 20.07 to 2552.41, with an average score of 78.85. A total of 2,787 contigs were identified to have a SNP present, representing 39% of the genome.

Dataset	Read sets	File size	Sequencing platform	Read type	Mapping results	No. of SNPs	Intersect
Hannover transcriptome	1	12.9MB	454	Single	96.43%	63,592	32,201
	2	13.5MB	454	Single	96.42%	69,440	

***Table 23**: Results from the first round of SNP calling using the GATK pipeline for identification of SNPs in D. gallinae. File size, sequencing platform, read type, mapping results and number of SNPs identified are labelled for both read sets from the Hannover transcriptomic dataset*

5.4.3 Intersected VCF files: Hannover read sets one and two

Intersection of VCF files demonstrated a total of 32,201 SNPs in common between both transcriptome read sets (Table 23). Intersecting SNPs were found in 1,914 contigs, representing ~27% of the genome. From these 1,914 contigs, 271 of them had a single SNP and the remaining 1,643 contigs had multiple SNPs identified.

Phred quality scores ranged from 20.07 to 2552.40, with the average of 120.30. Due to merging of VCF files, quality scores reflect those identified in read set two with a higher average demonstrated due to combining with read set one. More broadly, 74 SNPs had a quality score higher than 1,000; 733 SNPs had a score over 500 and 12,492 had a score over 100. The phred score (Q) is logarithmically related to error probability and can be calculated by the equation $Q = -10\log E$ (484). A phred score of 20 is taken as a reliable threshold, equivalent to a false-positive rate of 1% (456). Under this assumption all SNPs incorporated in the VCF intersect passed a reliable threshold. As such, the entire VCF was used to provide a reference file for the second round of SNP calling.

5.4.4 SNP discovery round 2

A total of 135,736 SNPs were called from both transcriptome read sets, of which 133,542 passed all filters applied (Table 24). For read set one; 65,248 SNPs passed variant filtration, with 1,048 failing the TruthSensitivtyTranche filter. Phred quality scores ranged from 30.00 to 3547.80, with an average of score of 114.09. A total of 2,486 contigs were identified to have a SNP present, representing 37% of the genome. For read set two; 68,294 SNPs passed variant filtration with 1,446 failing the TruthSensitivtyTranche filter. Phred quality scores ranged from 20.07 to 2552.41, with an average score of 78.59. A total of 2,787 contigs were identified to have a SNP present, representing 39% of the genome.

	GATK: Round 1 of SNP calling			GATK: Round 2 of SNP calling				
Dataset	Read sets	No. of SNPs	VCF intersect	No. of SNPs total	No. of SNPS PASS	No. of excluded SNPS	VCF intersect	VCF intersect PASS
Hannover transcriptome	1	63,592	32,201	66,296	65,248	1048	32,940	32,599
	2	69,440		69,440	68294	1146		

Table 24: *Results from the second round of SNP calling using the GATK pipeline to identify SNPs in D. gallinae. Information is provided for each transcriptomic read set for the number of SNPs identified from each round of SNP discovery. Columns outlining the number of SNPs which passed the filtering process and the number excluded are included for round two of SNP discovery, with the total number of SNPs from the intersect file shown and the SNPs which passed filtration.*

5.4.5 Intersected VCF files: Hannover read sets one and two Round 2

Intersection of VCF files demonstrated a total of 32,940 SNPs in common between transcriptome read sets, of which 32,599 passed all filters (Table 24). Out of these 32,599, when compared to the genome assembly, 62% of SNPs from the combined Hannover datasets were reference/alternative (20,119), carrying one copy of the REF and ALT alleles and 38% of SNPs were homozygous alternative (12,480). SNPs were found in 1,928 contigs, representing ~27% of the genome. Quality scores ranged from 20.07 to 2552.40, with the average 118.11. More broadly, 74 SNPs had a quality score higher than 1,000; 733 SNPs had a score over 500 and 12,571 had a score over 100. Filtered depth (read depth) ranged from 2 to 127, with an average of 11, and the phred-scaled likelihood of possible genotypes ranged from 0.06 to 99, with an average of 42.

5.4.6 Substitution type

Breaking down 32,599 SNPs in the dataset into substitution type, indicated that the frequency ranged from 3.85% (C-G) to 16.39% (G-A). The most common substitutions were A-G, C-T, G-A and T-C (Figure 29), with these four substitutions accounting for 63.66% of all SNPs discovered (Table 25, Figure 29).

Substitution type	Total number	% of dataset
A-C	1451	4.45
A-G	4999	15.33
A-T	1589	4.87
C-A	1563	4.79
C-G	1255	3.85
C-T	5203	15.96
G-A	5342	16.39
G-T	1654	5.07
G-C	1279	3.92
T-A	1617	4.96
T-C	5209	15.98
T-G	1438	4.41

Table 25: *Type, total number and percentage of substitution types in the final VCF intersect file, produced from intersection of the second round of SNP calling comparing D. gallinae transcriptomic read sets to the genome assembly*

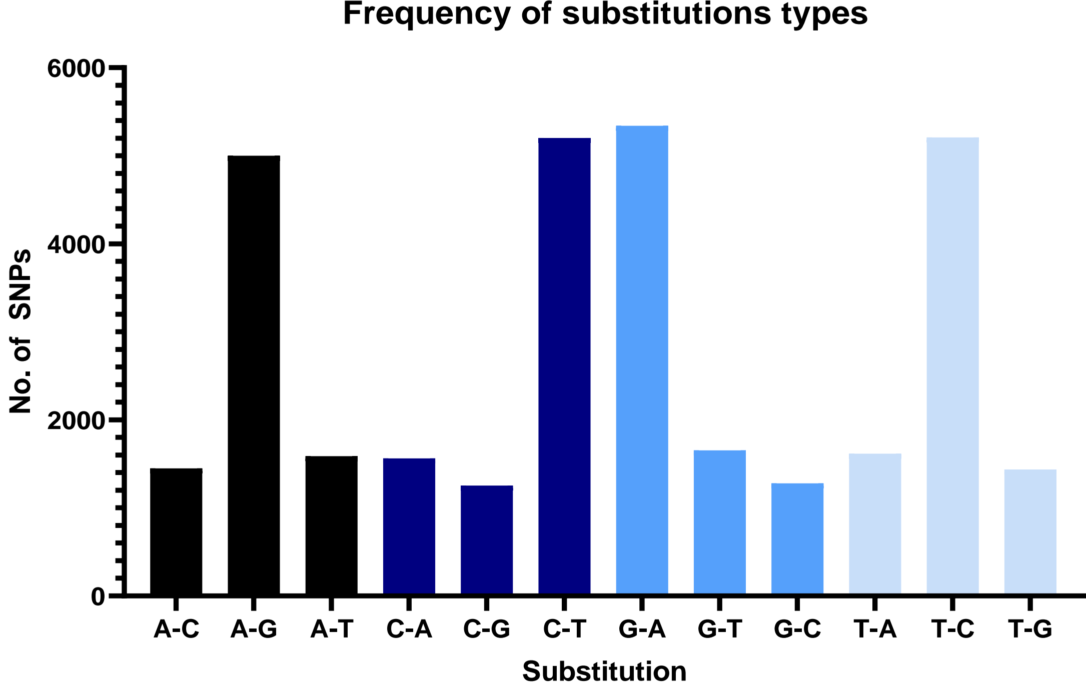


Figure 29: *Graph showing the number of SNPs for each substitution type in the VCF intersect file of D. gallinae transcriptomic read sets compared to the genome assembly. Black; SNPs from A-G/T/C, Dark blue; SNPs from C- G/T/A, Medium blue; SNPs from G- A/T/C, Light blue; SNPs from T- A/C/G*

5.4.7 Co-variate analysis: Comparison of RMSE values between rounds

5.4.7.1 CycleCovariate: Empirical data quality difference vs cycle

Comparison of pre- and post-table recalibration for the cycle covariate showed a reduction in root-mean-square deviation (RMSE) of 98.55% in round one for RMSE_good and 98.00% for RMSE_all (Figure 30). For round two, a reduction of 97.77% was observed for RMSE_good and 97.10% for RMSE_all (Figure 30).

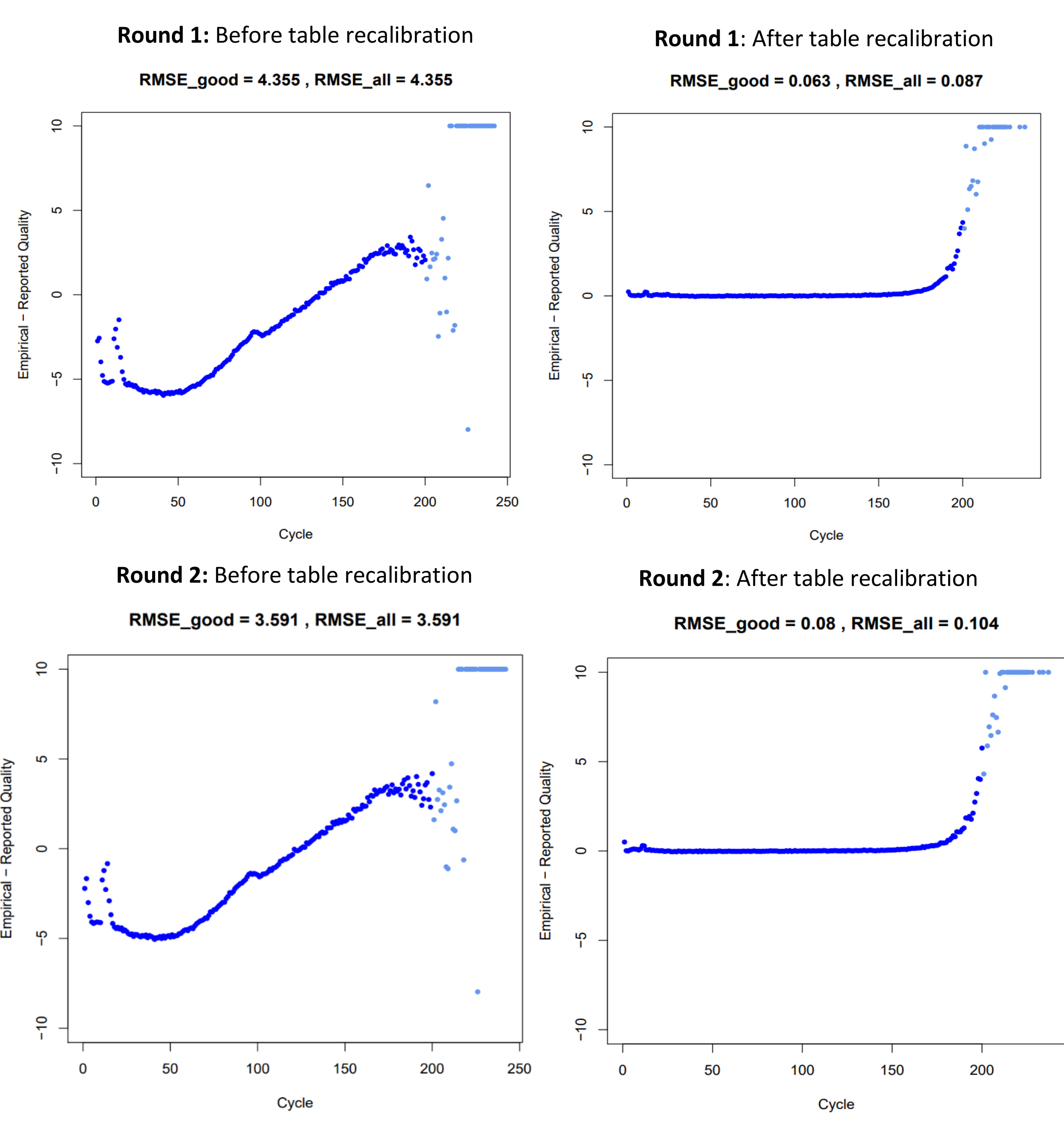


Figure 30: *Comparison of RMSE values between rounds of SNP discovery in D. gallinae for Co-variate analysis. CycleCovariate results for Hannover read set one; top row round 1 SNP calling, left before recalibration and right after recalibration, bottom row for second round of SNP calling, left before recalibration, right after recalibration. Similar results were observed for read set two so were not shown in duplicate*

5.4.7.2 DinucleotideCovarite: Empirical data quality vs cycle

Comparison of the dinucleotide covariate for pre- and post-table recalibration showed a reduction in RMSE of 98.5% in round one and 97.98% in round two (Figure 31). Comparatively between SNP calling rounds, RMSE was 17.84% lower in pre-table recalibration in round two than round one (Figure 31).

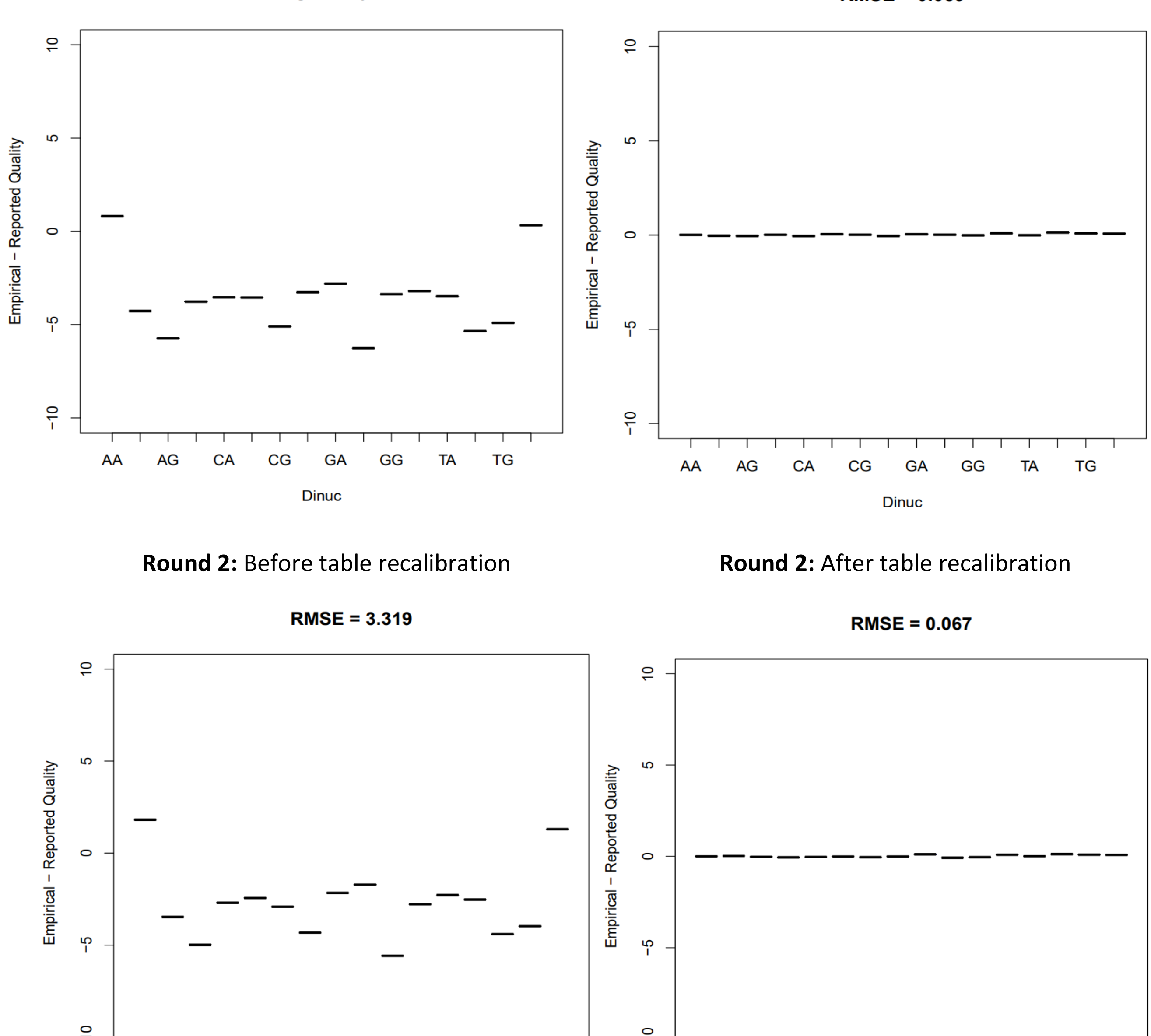


Figure 31: *Comparison of RMSE values between rounds of SNP discovery in D. gallinae for the Dinucleotide covariate. DinucleotideCovariate results for Hannover read set one; top row round 1 SNP calling, left before recalibration and right after recalibration, bottom row for second round of SNP calling, left before recalibration, right after recalibration. Similar results were observed for read set two so were not shown in duplicate.*

5.4.7.3 *QualityScoreCovariate: Empirical data score vs reported data score*

RMSE scores for the quality score covariate showed a reduction of 99.26% for RMSE_good and 99.20% for RMSE_all in the first round of SNP calling (Figure 7). For the second round of SNP calling a reduction of 99.20% was seen for RMSE_good and 99.40% for RMSE_all (Figure 32). Between rounds, before table recalibration a drop of 5.7% was seen in RMSE values.

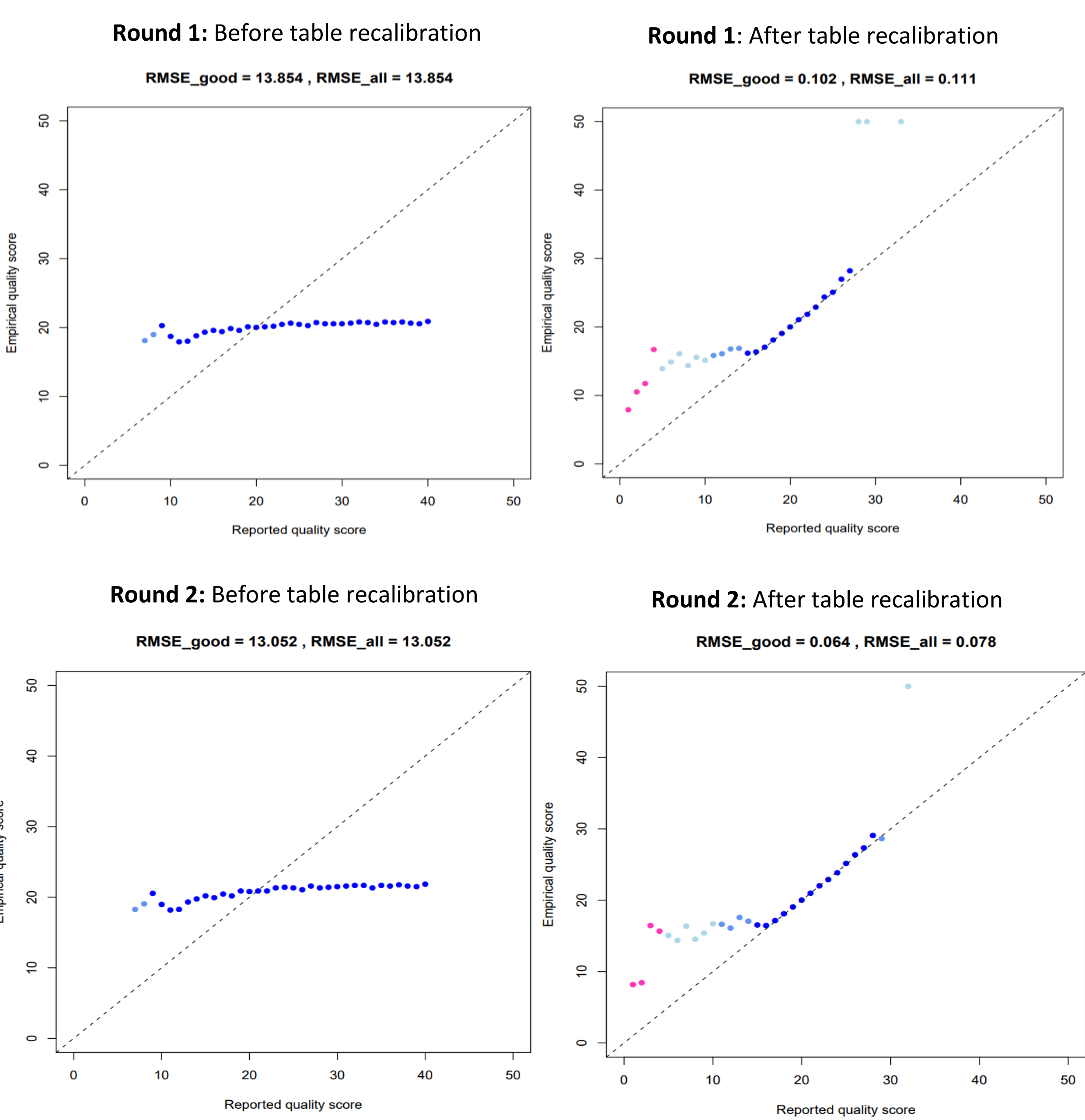


Figure 32*: Comparison of RMSE values between rounds of SNP discovery in D. gallinae for the quality score covariate. QualityScoreCovariate results for Hannover read set one; top row round 1 SNP calling, left before recalibration and right after recalibration, bottom row for second round of SNP calling, left before recalibration, right after recalibration. Pink plots refer to uncalibrated data and blue plots refer to calibrated data. Similar results were observed for read set two so were not shown in duplicate.*

5.4.7.4 Variant Recalibrator: Gaussian plots mixture model plots

The Gaussian mixture model reports for read set two are shown below (similar model outputs were observed for read set one). Gaussian mixture model reports were generated by VSQR for both read sets during round two of SNP discovery (Figures 33-35). In each report four panels display 2D projections of the model. The upper left panel displays the probability density function that was fit to the data, with green areas showing locations in the space indicative of high quality whilst red areas are indicative of lower quality. SNPs located in red regions were generally filtered out during VSQR (483). The remaining three panels in the mixture model plots provide scatter plots (with normalised units) whereby individual SNPs are plotted in the two annotation dimensions as points within a point cloud (483). The upper right panel displays which SNPs were retained (black) and filtered (red) through the VSQR procedure. The bottom left panel shows distribution of the SNPs used to train the model, with green representing SNPs found in training steps which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being true due to distance from learned Gaussians. Lastly, the bottom right panel displays SNPs by status (novel or known) to understand if the annotation dimensions show a clear separation between known SNPs and novel SNPs (483).

5.4.7.4.1 FisherStrand

Gaussian mixture model plots of homopolymer run (HRun), MappingQualityRankSumTest (MQRankSum), QualbyDepth (QD) and ReadPosRankSum against FisherStrand (FS) can be seen in Figure 33. Plotted against Hrun, the distribution of high quality SNPs can be observed clustered at the lower values for both FS and HRun and the distribution of negative training SNPs and filtered SNPs reflecting those of lower quality (i.e. in the red region of probability density function plot) (Figure 33), demonstrating ideal clustering. That is, a clustering showing that SNPs of poor quality were successfully filtered out during VSQR. Novel SNPs were found distributed in both high- and low-quality regions, with those located in high quality regions retained.

Mixture plots comparing MQRankSum and QD demonstrate similar distributions (Figure 33) with FS values remaining below ~4 but ranged from negative to positive for the compared annotation. In contrast to HRun vs FS, distribution of negative and positive training SNPs and filtered and retained SNPs showed admixture. This suggests that a portion of SNPs considered high quality

were filtered out during VSQR. Distribution of novel and known SNPs were mixed across both high- and low-quality regions.

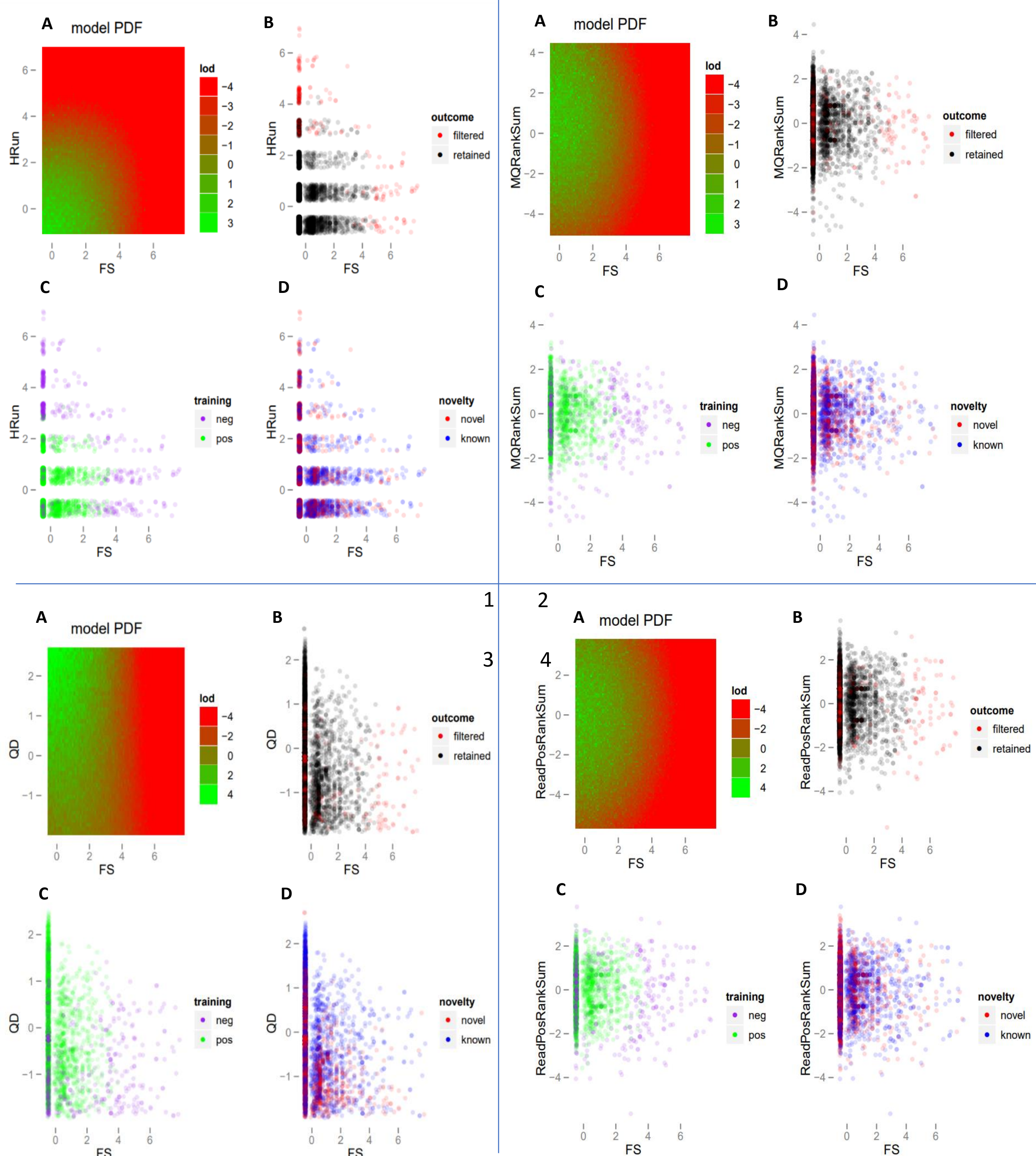


Figure 33: *Gaussian mixture model reports for HRun (1), MQRankSum (2), QD (3) and ReadPosRankSum (4) plotted against FS as part of the VSQR process. Produced on usegalaxy. Probability density function with green areas indicative of high quality and red areas indicative of lower quality displayed in panels 1-4 A. Retained SNPs (black) and filtered SNPs (red) shown in panels 1-4 B. Distribution of the SNPs used to train the model, with green representing SNPs which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being retained shown in panels 1-4 C. SNPs by status (novel or known) shown in panels 1-4 D.*

5.4.7.4.2 Homopolymer Run

Gaussian mixture model plots of MQRankSum, QD and ReadPosRankSum against HRun all displayed similar distributions (Figure 34). HRun values remained under 4 but other annotations ranged from negative to positive, however wider distribution was observed for QD vs HRun, with MQRankSum vs HRun and ReadPosRankSum vs Hrun displaying a curved distribution in favour of smaller values. The probability density function distribution was reflected in the scatter plots for all annotations, with distribution of negative and positive training SNPs and filtered and retained SNPs showing limited admixture.

This suggests that a small portion of SNPs considered high quality were filtered out during VSQR but that low quality SNPs were successfully filtered out in most cases. Distribution of novel and known SNPs were mixed across both high- and low-quality regions.

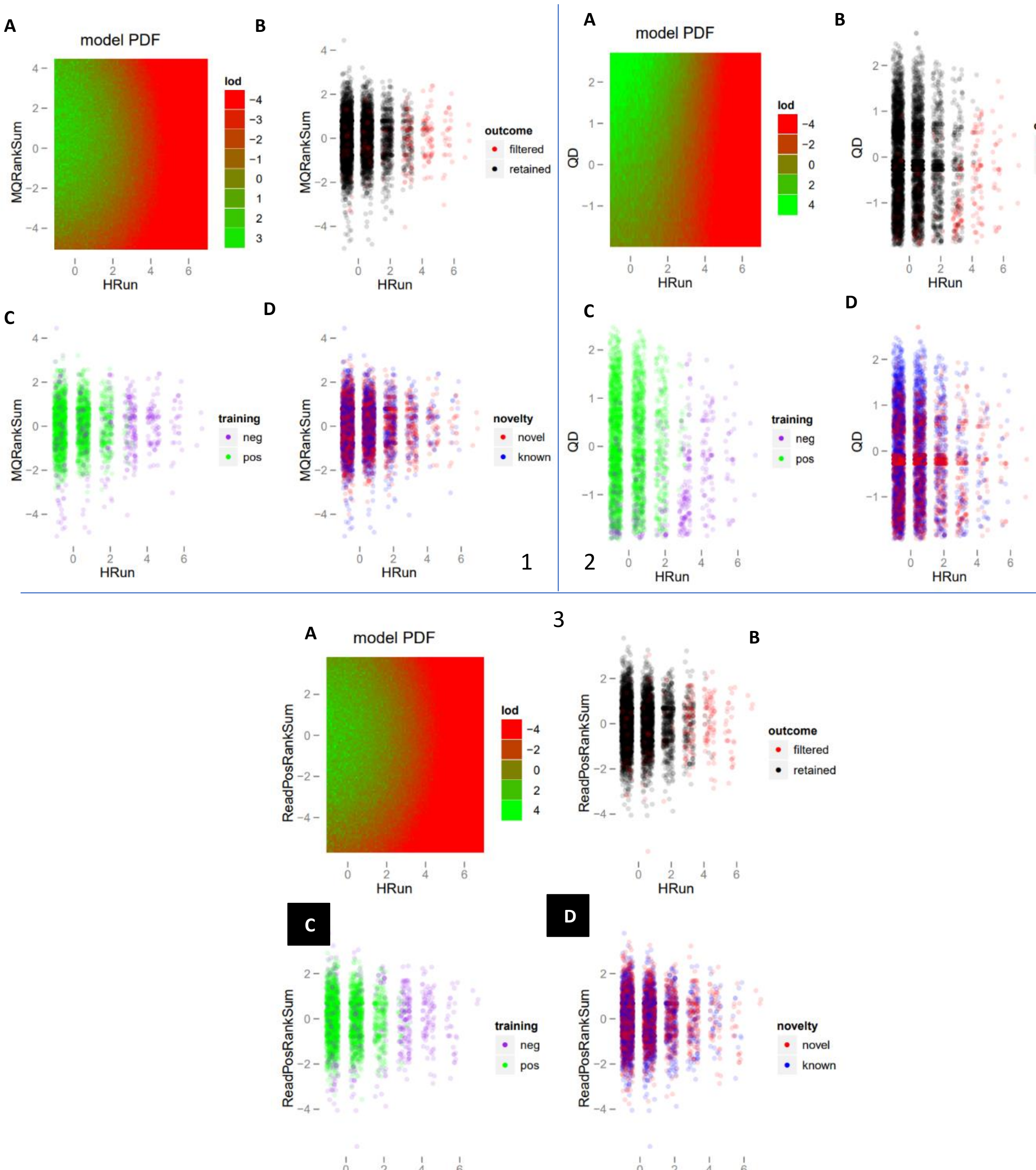


***Figure 34**: Gaussian mixture model reports for MQRankSum, QD and ReadPosRankSum plotted against HRun as part of the VSQR process. Produced on usegalaxy. Probability density function with green areas indicative of high quality and red areas indicative of lower quality displayed in panels 1-3 A. Retained SNPs (black) and filtered SNPs (red) shown in panels 1-3 B. Distribution of the SNPs used to train the model, with green representing SNPs which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being retained shown in panels 1-3 C. SNPs by status (novel or known) shown in panels 1-3 D.*

5.4.7.4.3 MQRankSum and QD

Gaussian mixture model plots of ReadPosRankSum and QD plotted against MQRankSum and ReadPosRankSum plotted against QD showed wide distributions, spanning negative to positive values compared to previous plots (Figure 33-34). Scatter plots revealed less separation between training (negative and positive points) and distribution less similar to the probability density function as in the other mixture model plots (Figure 35). A similar clustering of filtered and retained SNPs is observed for all annotation types (Figure 35).

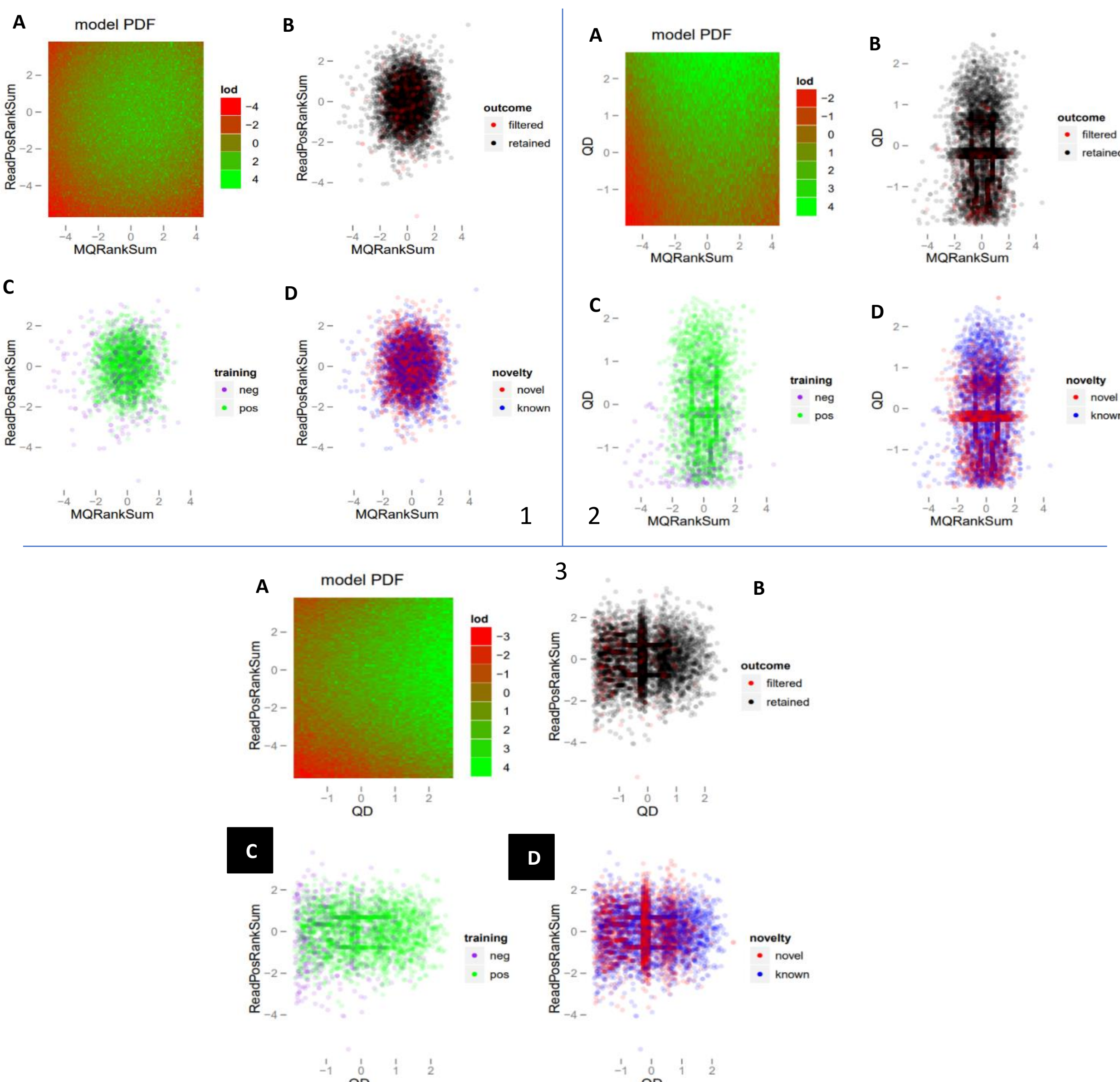


Figure 35: *Gaussian mixture model reports for ReadPosRankSum and QD plotted against MQRankSum, and ReadPosRankSum plotted against QD as part of the VSQR process. Produced on usegalaxy. Probability density function with green areas indicative of high quality and red areas indicative of lower quality displayed in panels 1-4 A. Retained SNPs (black) and filtered SNPs (red) shown in panels 1-4 B. Distribution of the SNPs used to train the model, with green representing SNPs which successfully passed the VariantRecalibrator step and purple representing SNPs with the lowest probability of being retained shown in panels 1-4 C. SNPs by status (novel or known) shown in panels 1-4 D.*

5.4.8 Read sets comparisons through each SNP calling round

When comparing genome coverage and quality score for read sets, genome coverage was higher for read set two for both rounds of SNP discovery (5% and 2%, respectively) but had a lower average quality score than read set one (Figure 36). There was no change in genome coverage recorded between rounds for read set one or for the intersected files. Genome coverage and quality score both increased in the second round of SNP calling for read set 1, rising 3% for genome coverage and 2.4% for quality score (Table 23, Table 24, Figure 36). Between rounds, there was no change in genome coverage in the intersected files and a slight reduction in average quality score (1.8%).

Intersection of VCF files from read sets resulted in a loss in genome coverage (Figure 13), of approximately 10% compared to read set one and 12% compared to read set two, in the second round of SNP calling. The average quality score for intersected files was comparable to read set one and higher than read set two (Figure 36).

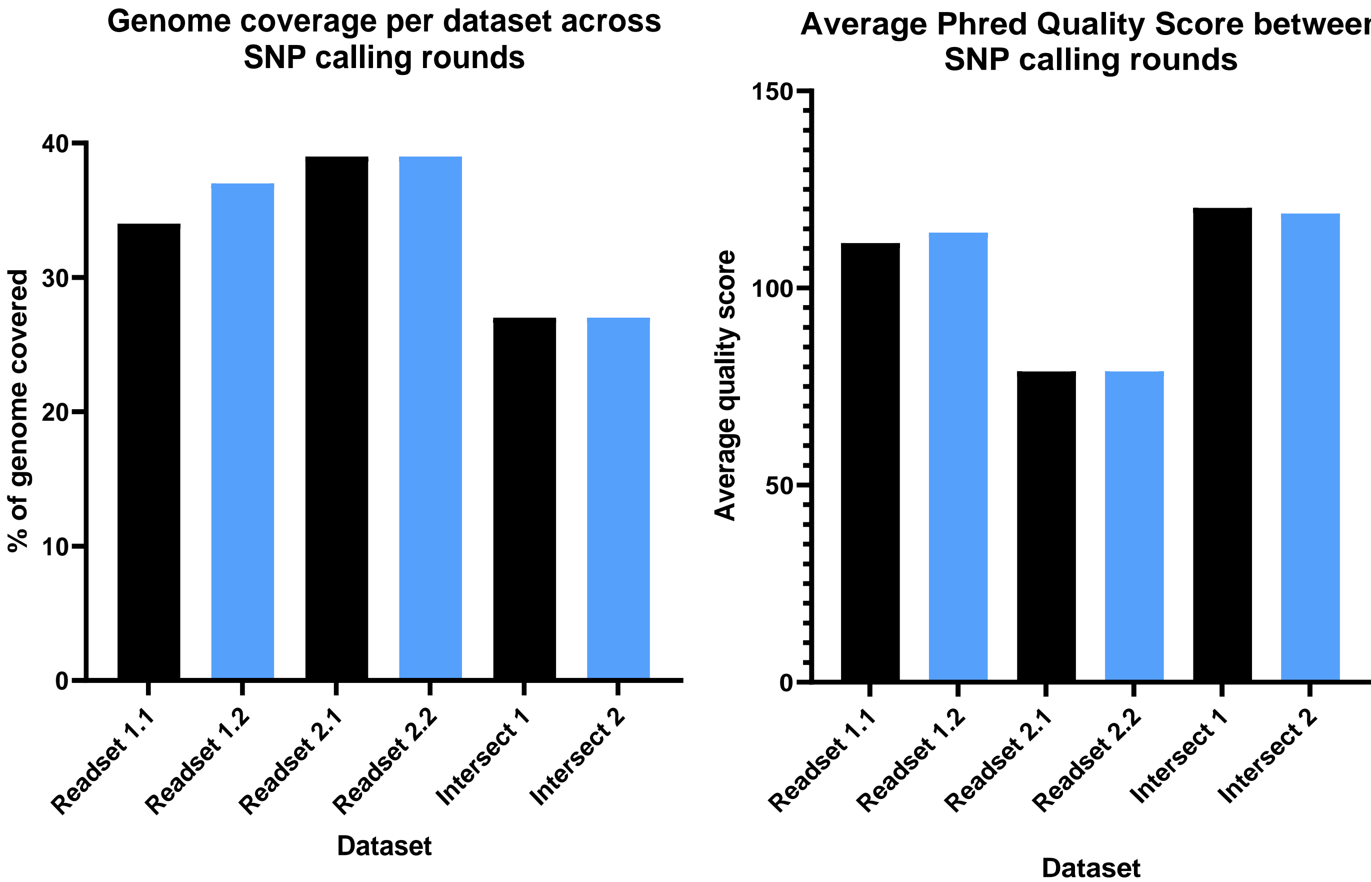


Figure 36: *Comparison of genome coverage and average Phred quality score for individual read sets and intersected VCF files for D. gallinae SNP identification using GATK. Left: Genome coverage (calculated by % of contigs from genome assembly with at least one SNP present) for read sets and intersected files for each round of SNP calling, Right: Average Phred quality score between SNP calling rounds. Round one indicated by black and round two indicated by blue colouration*

5.5 Discussion

Through utilisation and adjustment of the GATK best practices for germline SNP and indel discovery pipeline, a final SNP dataset of 32,599 high quality SNPs were identified when comparing transcriptomic data produced from *D. gallinae* collected from the University of Hannover (468) to the *D. gallinae* genome assembly produced from mites collected from a layer farm in Scotland (232).

5.5.1 SNP database for future studies

The SNP database generated provides a valuable set of candidate genetic markers to increase understanding of genetic variation in *D. gallinae*. Markers such as these are essential to allow SNP genotyping to be undertaken. There are multiple applications that this SNP database could aid in, including: genetic diagnostics (e.g. in pharmacogenomics (485)), genetic diversity analysis, phylogenetic analysis and characterisation of genetic resources (432, 438). A total of 32,599 SNPs, 62% identified as reference and alternative and 38% as reference or alternative, are available to be used and selected from for various downstream applications. The average Phred score for the dataset set was 118.11, with a minimum score of 20 (equal to 1% false-positive rate) for all SNPs, typically considered a reliable threshold (456). Overall, the SNPs identified here were restricted to the exome given use of transcriptomic datasets. Comparison with the draft genome assembly found that ~27% of the contigs from the *D. gallinae* genome assembly contained at least one SNP. Due to the unannotated nature of the genome assembly it is currently difficult to provide a deeper description of which genes the SNPs might be located in, or any links with gene function.

5.5.2 Biological validation

After generation of a list of filtered SNPs, it was important to provide confirmation through validation using an alternative approach (486). One way of achieving this is through the design of primers flanking the SNP, and Sanger sequencing of the resulting amplicon, or alternatively high-throughput SNP genotyping methods including SNP assays (487, 488), high-resolution melting (HRM) (489), mass spectrometry (490) or amplicon sequencing (491). Biological validation of this *D. gallinae* dataset was validated using Sanger sequencing and will be discussed in the next chapter 'SNP genotyping of *D. gallinae* through a next generation sequencing multiplex platform'.

5.5.3 GATK pipeline complications and solutions for *D. gallinae*

The GATK toolkit was originally designed for the analysis of the human genome and exome before being expanded to handle other organisms (492). At the time of completing the pipeline (2017-2019), a lack of information surrounding SNPs in *D. gallinae* genomic and transcriptomic data, being largely unassembled and unannotated, led to complications in completing the best practices for germline SNP and Indel variant discovery. Complications arising from a lack of pre-existing data on known variants in the Base Quality Score Recalibration (BSQR) procedure step of the pipeline resulted in the decision to adapt the best practices workflow to form a self-validating pipeline. Evidence that this was successful is demonstrated when comparing results from the initial and second round of SNP discovery. The second round of SNP discovery yielded 1,656 additional SNPs (that passed variant filtration) for read set 1, compared to the first round (Table 24). For the second read set, the exact same number of SNPs was called in both SNP discovery rounds, however 1,146 were discarded during variant filtration in round two (Table 34).

Comparing co-variate analysis plots from pre- and post-table recalibration shows a clear reduction in RMSE in both rounds (Figures 30-33). Table recalibration produced RMSE values lower in round one for the cycle covariate and dinucleotide covariate, however there was a difference of 0.55% in overall reduction for the CycleCovariate (Figure 7) and 0.52% for the dinucleotide covariate (Figure 30). Focusing on the quality score covariate, reductions of 99.26% and 99.20% RMSE were observed when comparing pre- and post-table calibration and a lower RMSE score was seen at the end of base recalibration for round two. In all cases, it can be observed that RMSE values pre-calibration were lower in the second round of SNP discovery, demonstrating a positive impact of the addition of the VCF file as a reference-ordered file (produced from round one) (Figures 30-33). Comparing the average quality score of read sets between SNP discovery rounds showed no change for read set and an increase in quality for read set one in the second round (Figure 36).

Gaussian mixture model plots from the VSQR process demonstrated mixed clustering results but, in principle, provide further support that the self-validation strategy worked. The best clustering of SNPs was seen in plots for HRun vs FS (Figure 33), with clustering of SNPs at low values for each of the statistics and filtration of SNP points located in red regions of the probability density function plot indicative of good clustering (483). Thus, SNPs of low quality were filtered, and SNPs of high quality retained, providing support that addition of ‘known’ SNPs in the second round of calling helped to reduce the likelihood of false SNP discovery through VSQR. Gaussian mixture model plots comparing MQRankSum, QD and ReadPosRankSum to FS

or HRun, respectively, clustered so that the main statistic (FS or HRun) retained a low value but the other statistic ranged from negative to positive. Despite this, the scatter plots demonstrate that most SNPs located in the red regions of the probability density plot were filtered out during VSQR, demonstrating support for the self-validating pipeline further. In three plots (Figure 35) there was crossover between SNPs filtered and retained, indicating the possibility that a number of low-quality SNPs were retained or a number of high quality SNPs were filtered unintentionally. It should be noted that the distribution observed in the probability density function plot for these models were wide, with distribution spread across almost all values.

5.5.4 SNP identification in related species

Studies focusing on SNP identification through similar pipelines such as GATK in acari or related species are quite sparse in the literature. Current studies available utilise a range of DNA extraction techniques, number of isolates used, sequencing platform and bioinformatics tools and pipelines for SNP identification. This makes a direct comparison difficult, in addition some studies focus on identification of selected SNPs, e.g. those relating to resistance. Genome size and annotation status also varies depending on the species, further complicating comparisons.

Work focusing on spider genomes to provide an insight into composition and evolution of venom and silk involved the genome assembly and annotation of the African social velvet spider, *Stegodyphus mimosarum* and the mygalomorph Brazilian white-knee tarantula, *Acanthoscurria geniculata*. They generated a *de novo* assembly for the velvet spider from 91 x coverage sequencing of paired end and mate pair libraries, estimating genome size at 2.55 Gb. In total they called 1.09 million high-quality SNPs, which they achieved through mapping back to scaffolds. In the Brazilian white-knee tarantula a total of ~2.2 million SNPs were identified from an estimated genome assembly size of 6.5Gb (493). The estimated genome size of both these species is considerably larger than *D. gallinae,* which has a genome that is 959 Mb in size. Analysis completed in the study utilised genomes rather than transcriptomes, therefore factoring this in it is anticipated that a greater number of SNPs would be identified.

One study focusing on *Apis mellifera* (Hymenoptera: Apidae) utilised a very similar GATK pipeline to aid in development of SNP markers specific to a line displaying high hygienic behaviour against *Varroa destructor*. They extracted genomic DNA from 40 individual honeybees, 20 showing high hygienic behaviour (HHB) and 20 from low hygienic behaviour (LHB) lines. They produced a total of 1.3 million reads using Illumnia NextSeq500 sequencing platform which they mapped and aligned to the genome (Amel 4.6). The Picard toolkit was used for removal of potential duplicates and variant calling was completed using the same 'UnifiedGenotyper' and 'SelectVariants' GATK tools included in the *D. gallinae* pipeline. In total, after variant filtration,

~2.3 million SNPs were found in the HHB line and 3.2 million SNPs in the LHB. Interestingly, they note that the majority of SNPs were found in introns (27.5% in HHB and 28.3% in LHB) and intergenic regions (65.7% in HHB and 65.6% in LHB). They demonstrated identification of considerably fewer SNPS in genic regions, including UTRs, splice sites, transcripts and exons, with just 6.75% of SNPs in HHB and 6.1% of SNPs in LHB lines found in genic regions (494). Whilst the *A. mellifera* is not a closely related species to *D. gallinae,* this study highlights that a greater number of SNPs could be identified if comparison of whole genomes was completed, rather than comparison of transcriptomic data with genomic data. Comparison of entire genomes from multiple individuals would locate SNPs in introns and intergenic regions that have been missed in the current pipeline and could vastly expand the available dataset of SNPs, however the stability of these SNPs might be less robust than those identified from coding regions.

5.5.4.1 SNP identification in acari species

Research looking into the use of DISCOSNP, a computational method for identification of SNPs directly from sequence reads, without a reference genome, conducted a population genetics study of the castor bean tick, *Ixodes ricinus*. DNA was extracted from two tick pools, one from Gardouch, France (ten individuals) and one from Malville, France (twenty individuals). They applied a genomic reduction on the two pools corresponding to 3.8% of the initial genome. They generated a total of 996,508 reads (536,061 for the first pool and 460,447 for the second) with an average length of 529bp. Using DISCOSNP they identified 321,088 SNPs, of which they selected 384 to experimentally validate using Fluidigm technology. A total of 464 individuals were genotyped with 95.8% of SNPs displaying minor allele frequency varying between 0.04-0.5 (mean of 0.23). From the remaning SNPs (16 in total), five SNPs failed to amplify and 11 demonstrated only one of the two alleles (495, 496). A similar study focused on the population and evolutionary genomics of the lone star tick, *Amblyomma americanum*, using high-throughput genotyping-by-sequencing approach in 90 ticks from five locations (497). Extracted DNA was genotyped-by-sequence (GBS) to produce a demultiplexed and quality filtered dataset containing 235,335,191 reads from two lanes of sequencing. They utilised the UNEAK bioinformatics pipeline for alignment of raw sequence reads and calling SNPs. UNEAK is a multisample SNP calling method that was developed for analysis of GBS data from species which lack a reference genome (498). In total they found 72,517 biallelic SNPs (497). Another study on *A. americanum* looked at 189 ticks from across the species geographic range. They used a double digestion site-as-associated DNA sequencing technique to identify 81,818 SNPs (499). These studies produced a comparable number of SNPs to the GATK pipeline used for *D. gallinae* when factoring in they were comparing whole genomes.

In mite species, work on the two-spotted spider mite, *Tetranychus urticae* (Acari: Tetranychidae) revealed ~590,400 high quality SNP variants identified as segregating in the experimental population. They focused on multiple strains of *T. urticae* as the primary focus of the study was high-resolution QTL mapping to reveal acaricide-specific response and common target-site resistance after undergoing selection by different Mitochondrial Electron Transport Inhibitors of complex I (METHI-I) acaricides. They extracted genomic DNA from 4 x 200 adult mites/population that they pooled and precipitated together. They generated Illumina genomic DNA libraries producing paired-end reads of 101bp-1250bp, which they aligned to the reference draft *T. urticae* genome using the Burrows-Wheeler aligner. Similar to the pipeline in this study, they followed GATK best practices, using the UnifiedGenotyper tool to produce the final VCF file. General challenges involved in SNP discovery

Differentiating machine artefacts, resulting from the high rate and context-dependant nature of sequencing errors, from true genetic variation is the outstanding challenge involved in analysis of NGS data (472). Three of the challenges faced during SNP discovery include sequencing platform error rates, sequencing read length and correct mapping of reads to a reference genome (222, 227, 230, 462, 472, 500-503) .

5.5.4.2 Sequencing platform error rates

One of the challenges of *in silico* SNP discovery is not identifying polymorphic sites but differentiating these from sequencing errors (227, 472, 500). This is particularly true when considering NGS compared to traditional DNA sequencing, due to a higher error rate. Next-generation sequencing technologies are prone to inaccuracies as often as one every 20 bp, but this rate varies substantially between technologies and platforms and improves over time as companies continually strive for an improvement in accuracy whilst increasing read length. Electronic mining to identify polymorphisms can be impeded by the presence of read errors (500). Several different error types need to be considered when trying to distinguish true polymorphic sites from sequencing errors and the approach depends on which platform was used in the generation of the sequence data, owing to the fact each technology has a distinct error profile (227, 500).

The transcriptome produced by Schicht et al. (468) was produced through Roche 454 sequencing, which is known to have difficulties in quantification of homopolymers, resulting in insertions and/or deletions, and has an error rate of ~1% (463). Base calling for the 454 platform involves inference of homopolymer length from the observed fluorescence intensity and as such the main challenge faced is the large variance of signal intensity for specific homopolymers, causing the issue with high error rate regarding indel calls (227). For this study only SNPs were

selected and indels were filtered out, meaning that the variance faced by 454 calling indels was not a major issue for SNP discovery. The substitution error rate for conventional NGS technologies has been reported in the literature to be >0.1% by multiple studies (504-506). In context, this equates to 32.6 SNPs from the dataset produced from the GATK pipeline to have substitution rates.

5.5.4.3 Sequencing read lengths

One source of error occurs due to the short reads produced by NGS technologies, resulting in incorrect mapping to the reference genome. Occurrence of this error type is possible at any genomic region that has one or more similar copies elsewhere in the genome, for example due to polyploidy, multigene families or genome duplications (500). A major source of sequence errors stems from the balance between the drive to obtain the longest sequence length and the confidence at which bases are correctly called. As a result, trimming, filtering and processing of sequences is often undertaken to reduce the frequency of erroneous sequences (500). The sequencing of the available genome or transcriptomes were not conducted as part of this PhD project and, as such, sequencing technologies used for production were outside of the control of this study. The genome was constructed using a mixture of both short and long read sequencing technologies, PacBio reads and MinION. In part, incorporation of both short and long read technologies helps in overcoming the problems faced by either technology and pipelines such as GATK have been designed to take such errors into account.

5.5.4.4 Correct mapping to the reference genome

Mapping reads to a reference genome (222, 230, 462, 501-503) is one of the first and most critical computational challenges for SNP discovery, requiring that every read must be aligned independently, and thus guarantees multiple reads which span indels will be misaligned (472). Per-base quality scores are used to illustrate what possibility a called base in a read is a true sequence base (507) but are not always accurate and vary depending on sequence context, machine cycle and sequencing technology used (508-510). Downstream, inaccurate quality scores and misaligned reads will propagate into the SNP discovery process and into genotyping. This is a general issue but typically becomes acute in projects that involve multiple sequencing technologies coming from multiple centres that rely on rapidly changing experimental pipelines, including the 1000 genomes project (472). Even once well-mapped, calibrated and aligned reads are produced, resolving SNP's or more complex forms of variation, such as indels, copy number variations or multi-nucleotide substitutions, still depend on specific and sensitive statistical models (462, 473, 475, 501-503, 510-518). In addition to the selection of an appropriate alignment software, three issues are worthy of consideration. Firstly, paired-end reads have

been proven to be a viable solution for overcoming the ambiguity surrounding mapping short reads to a reference genome (519) . Secondly, reads that can be mapped but result in numerous mismatches should be eradicated and as such any mutations which are only supported by these reads should not be included in downstream analysis. Finally, as many NGS technologies incorporate PCR steps into part of their library preparation process, multiple reads that originate from one template can be sequenced, which can cause interference in variant calling statistics. As a result, it is thought of as common practice to remove PCR duplicates after alignment in both whole-genome and whole-exome research (519).

These factors were considered during pipeline development. BWA-MEM mapping was chosen to be an appropriate alignment software, as it includes an alignment algorithm designed for the alignment of sequence reads against large reference genomes (474). Overall BWA-MEM mapping showed 96.4% mapping of both transcriptome read sets to the genome (Table 3). In future, additional tools could be used for filtration and removal of low-quality reads to further reduce errors. The complication arising from PCR duplicates was dealt with during pre-processing (phase one) through utilisation of the MarkDuplicates step. In conjunction, comparison of SNPs identified with other transcriptomes available (RVC and Moredun) would provide further support for the validity of SNPs discovered through the GATK pipeline. Manual comparison with RVC and Moredun transcriptomes was conducted through blasting of fragments (corresponding to selected SNPs (outlined in the following chapter) but due to time constraints and complications, the RVC and Moredun transcriptome were not put through the GATK pipeline.

5.5.5 Future considerations

5.5.5.1 Using a consensus approach

For variant identification Pabinger et al., recommend using a consensus approach, e.g. running multiple tools on the same dataset, such as CRISP, GATK and SAMtools. In the literature, it is has been reported that a single approach is incapable of comprehensively capturing all genetic variation present, thus, applying multiple variant identification tools could be the most successful approach (519). Variants could then be filtered only if they fulfil specific criteria, for example being identified by a minimum level of variant identification tools (520). The downside to applying such a strict criterion is the filtering out of true positives (519).

In an ideal situation, running multiple pipelines for the *D. gallinae* data would have been beneficial. However, the same complications on the lack of a reference SNP database and time consuming processing in the GATK pipeline (discussed in 2.3 Methodology: subsections

5.3.3.3.1.1.1, Complications with *D. gallinae* dataset, and5.3.3.3.1.1.2 Solution to lack of known variants) would apply to other pipelines that resulted in significant time delays in SNP discovery. It was not feasible to run multiple variant discovery pipelines due to the time constrain of the project. One time constraint was that the genome assembly proved critical for successful completion of the GATK pipeline and was not available until late 2018.

5.5.6 Conclusions

Overall, use of the GATK pipeline for germline SNP and indel variant discovery resulted in identification of 32,599 SNPs, covering 27% of the genome assembly contigs, noting that variants were restricted to exonic sequences given the use of transcriptomic sequence data for analysis. Complications with base quality score recalibration, due to lack of pre-existing data, were resolved by implementing a self-validating pipeline. This SNP database offers a solid foundation to allow for a selection of targeted SNPs for SNP genotyping, aiding in a greater understanding of the genetic diversity of *D. gallinae*. The hypothesis that *D. gallinae* transcriptomic resources could be utilised in identification of SNPs which aid in defining the *D. gallinae* genome can be accepted.

6 GENOME-WIDE GENETIC ANALYSIS: SNP GENOTYPING OF *D. GALLINAE* USING AN NGS MULTIPLEX PLATFORM

6.1 POPULATION GENETICS

Population genetics can be defined as the study of the distribution in time and space of allele frequencies, or patterns, resulting from evolutionary processes and forces (428, 521, 522). Characterisation of a population's allele distributions and frequencies enables inferences to be made regarding processes (mutation, gene flow, natural selection and genetic drift) that have contributed to shaping patterns observed for a certain population (e.g. divergence, clustering, differentiation etc.) (523). A central goal of population genetics has been determining the evolutionary forces that affect genetic variation (524). Statistical inference can be used to estimate population genetic parameters, such as genetic diversity, effective population size, population differentiation, or phylogenetic relationships, and these population genetic metrics reflect processes that affect the genome as a whole (525). For example, genetic variation and population differentiation often vary tremendously across the genome due to variation in recombination rate, selection intensity (purifying and positive), and mutation rate (526).

6.1.1 Population genomics

Population genomics can be thought of as population genetics writ large, that is population genetic analyses conducted on a large number of loci which are distributed throughout the genome (527-529). Population genomics can be narrowly defined as separation of locus-specific effects (e.g. selection, recombination, mutation) that impact one or a few loci at a time from genome-wide demographic effects, including founder effects, genetic bottlenecks, inbreeding etc. (530). Through utilisation of a large number of loci spread across a genome, a comparison of the effects of selection on beneficial polymorphisms and neutral polymorphisms present at flanking sites (known as genetic hitch-hiking; (531)) can be undertaken against genome-wide demographic effects, that are not locus specific (530). This means that a population genomic approach can be described in four phases (528): (i) sampling of a large number of individuals, (ii) genotyping of this large population at multiple independent loci, (iii) identification of statistical 'outlier' loci and (iv) either estimation of demographic parameters and statistics (e.g. F_{ST} and phylogeographic structure,) in a large data set where outlier loci have been removed, or alternatively, focusing on the outlier(s) specifically in an attempt to infer potential selective mechanisms underlying them (530).

6.1.2 Episodes of selection

Previous episodes of selection can lead to distinct signatures in a population's genome (532). Balancing selection results in an excess of genetic variation in the region surrounding a locus under selection (533, 534), whilst directional selection causes harbouring genomic regions to

show reduced diversity (535), a local increase in linkage disequilibrium (536-538), and an alteration in allele frequency (539). A number of methods have been developed in order to detect loci under selection using patterns of genetic diversity observed within a population (348, 350, 401, 536, 540-543). However, selection pressures can also impact the genetic diversity between populations, as a locus experiencing balancing selection will show allele frequencies that are even across populations, whilst loci under directional selection can present with large differences between populations (544-546). In non-model organisms, global and pairwise F_{st} values are typically estimated over all loci; as all markers are assumed to be effectively neutral, there should not be any major inconsistencies between loci. However, when loci are potentially under different selective pressures the estimates may be different for each locus, requiring per locus estimates (547).

Population analysis should consist of asking one or more relevant biological questions, sampling representative individuals from the target population, determining allele frequencies at loci and using statistical approaches for inference of patterns and processes (522). A typical pipeline for a population genomic study has a number of crucial steps: (i) sequencing strategy design; (ii) sequence data generation; (iii) mapping of sequence reads; (iv) variant calling and genotyping; and (v) population genetic and/or molecular evolutionary analysis (548). For this study, steps one and two were conducted outside of this PhD thesis, while mapping and identification of SNPs has been discussed in Chapter four. Genotyping and population genetic analysis is the focus of this chapter.

6.1.3 Population Structure and Phylogeography

Populations exist across space, and spatial distribution is an important focus of population genetics (525). Quantifying population structure and levels of genetic differentiation among populations has been achieved with traditional population tools, but genomic techniques can provide greater statistical power and precision for estimating these parameters (549). Many analytical tools are well-suited for assessing and visualising population structure from large genome SNP datasets, such as principal components analysis and Bayesian clustering methods, as well as application of multiple techniques to a single dataset, which can aid in revealing important patterns (525). When applied to genome-wide data, these approaches help illustrate the results of processes that cause change across the whole genome, including population size and migration rates (525).

6.1.4 Estimating allele frequencies

A common goal of most population genomic studies is to either genotype each individual at variant sites or alternatively (and more commonly) use pooled population-wide data to directly

estimate allele frequencies (e.g. (550, 551)). It is possible to estimate genotypes and allele frequencies from the GATK/SAM tools output described above, but it has been shown that for low-medium coverage sites, this might introduce biases (552). Thus, alternative approaches have been developed using maximum-likelihood approaches to directly estimate genotypes from the sequences, without first calling SNPs (553, 554). Similarly, bias can also be introduced when calculating allele frequencies from low-coverage genotype data, for example due to loss of low-frequency alleles which can affect the site-frequency spectrum (555). Also in these cases, it seems appropriate to estimate allele frequencies directly from the sequence data, using alternative statistical approaches (e.g. (556, 557).

6.1.5 SNP genotyping

SNP genotyping is rapidly becoming a powerful tool for assessing genetic variation in natural populations (427, 558, 559). Recent applications show in some species that SNPs are extremely common and can be straightforward to ascertain in many non-model organism's genomes (560-562). Such SNPs can be used in a wide range of population studies, from individual identification to population structure and taxonomy (562-566). Benefits of using SNPs compared to other nuclear markers such as microsatellites include the ease and efficiency of discovery and genotyping (e.g. (562)), ability to target variation in either known genes or random genomic regions (561, 567), and existence of theoretical treatment (568-572) and analysis tools and/or methods that can assess power and population parameters (573, 574). Additionally, high-throughput genotyping has been shown to improve results for poor quality samples (such as historical or degraded samples (559, 575)), and provides the ability to examine both neutral variation and regions under selection (547). Despite microsatellites typically demonstrating greater allelic diversity per locus, individual SNPs can segregate reliably among populations (576, 577). The statistical power to detect population structure can be related to the total number of alleles examined, with the discriminatory power of approximately 100 neutral SNPs being roughly equivalent to approximately 10-20 microsatellite markers (571).

6.1.5.1 SNP genotyping in non-model organisms

The use of high-throughput SNP panels for study of many non-model organisms has primarily been limited by the cost and difficulties of identifying and studying new SNPs (i.e. in species lacking a curated genome and/or genomic resources), and consequently, the number of available assays has been low or non-existent for many species (437). However, technological advances and innovative methodologies are now enabling rapid SNP discovery (578, 579). With decreasing technology costs (580), SNP discovery projects are becoming more common, and the

number of novel SNPs potentially available for conversion to high-throughput assays is rapidly growing (for example (577, 579) and many others) (437).

Population studies in non-model organisms which have utilised high throughput assays for SNPs typically follow an initial discovery phase where every available marker was used (e.g. (581, 582)). Increasingly, researchers have become interested in developing SNP panels that are tailored to their specific research objective (583) and study system (581). Panels of SNPs can be developed and optimised for laboratory performance (i.e. to produce genotypes which are easily distinguished and reproducible), for specific genotyping platforms, and for power to resolve population structure (547, 584). Additional locus selection programs including WHICHLOCI (585) and BELS (586) can be used to rank and evaluate loci based on their performance for individual assignment and in certain instances mixed stock analysis (BELS) (587). However, there are complications involving upward bias in SNPs ranked in high-resolution loci when using these programs (588). At present there is no consensus on how to rank molecular markers, especially SNPs, however ranking and evaluating SNPs for use in a SNP panel is becoming increasingly important as the number of high-throughput assays continues to expand (437).

6.1.5.2 Prioritization of candidate variants

One of the challenges of next-generation genetics is narrowing down the list of candidate variants to a select panel and interpretation of the remaining variants within a biological context (589). One widely used approach for candidate list reduction is to include known variants which are present in published studies, public SNP databases, or in-house databases (590). Another approach is the use of pedigree information, i.e. sequencing distantly related individuals that have a phenotype of interest in order to identify the associated mutation(s) (591). This approach can also be helpful in the identification of the cause of common disorders which are genetically highly heterogeneous (519). After variant calls are generated, researchers need to gain an understanding of the descriptive content within the data and thus perform prioritization analysis on all variants, considering whether markers are phylogenetically informative, with functional follow-up on selected variants (219).

6.1.5.3 Variant interpretation

The process of converting quality scores and base calls into a set of genotypes for every individual in a sample can be divided into two steps: SNP calling and genotype calling. The aim of SNP calling, also known as variant calling, is to determine polymorphic positions, that is positions with at least one variable base compared to a reference genome or sequence (227). Variant interpretation involves consideration of the impact of a variant on a transcript or

protein, therefore it is dependent on protein or transcript identification and localising variants to whether they are protein-coding or non-coding regions (461).

6.2 AIMS AND HYPOTHESES

6.2.1 Aim of the study

The main aim of this study was to assess the population structure of *D. gallinae* using a Mid-Plex genotyping assay for genome-wide genetic analysis of UK and European *D. gallinae* populations. One objective required to achieve this was selection of a panel of high-quality SNPs from the 32,599 SNPs identified previously through the GATK piepline in the SNP genotyping chapter.

6.2.2 Hypotheses

Hypothesis 6: Diversity of *D. gallinae* will be greater between countries than within individual countries.

Hypothesis 7*:* There will be no significant difference in genome-wide diversity between different layer husbandry systems.

Hypothesis 8 Genetic diversity of *D. gallinae* populations will remain constant over time within farms.

6.3 METHODS

6.3.1 Sample selection

A total of 750 samples were collected for SNP genotyping. Originally, ten individual *D. gallinae* were isolated from every UK farm (see Table 4: Chapter three), totalling 240 samples, supplemented by five individual *D. gallinae* from every European farm (see Table 6: Chapter three), totalling 400 samples. The remaining 110 samples were selected from the same farm over time (Table 5: Chapter three), with five individual *D. gallinae* for UK7 and UK11 and ten individual *D. gallinae* for UK6 per timepoint.

Due to complications achieving the DNA concentrations required for Eurofins Mid-Plex genotyping from individual *D. gallinae* extractions, and the subsequent failure of some DNA extractions to pass Eurofins quality control, individual mite samples were eventually pooled per farm across the UK (Table 26) and Europe (Table 27) (see Tables 4-7, Chapter three); 50-80 mites per pool.

Country	County	Production type	Organic Status	Farm no.	Time points
England	Durham	Free-range	Non-organic	UK1	
	East Sussex	Free-range	Organic	UK 6	+0+14, +15,+20
	Gloucestershire	Free-range	Non-organic	UK2	
	Hampshire	Free-range*	Non-organic	UK24	
	Kent	Free-range	Non-organic	UK5	
	Lincolnshire	Free-range	Non-organic	UK14	
	Oxfordshire	Free-range	Non-organic	UK7	+0,+2, +14
	Suffolk	Free-range	Non-organic	UK11	+0, +10, +15
	West Sussex	Free-range	Non-organic	UK12	
	Shropshire	Intensive	Non-organic	UK8	
	Tyne and Wear	Intensive	Non-organic	UK15	
	N/A	N/A	N/A	UK18	
Northern Ireland	Tyrone	Free-range	Non-organic	UK10	
	Tyrone	Free-range	Non-organic	UK22	
	Tyrone	Free-range	Organic	UK23	
Scotland	Peebleshire	Free-Range	Non-organic	UK20	
	Highlands	Free-range	Organic	UK9	
	Peebleshire	Intensive	Non-organic	UK13	
Wales	Cardiganshire	Free-range	Organic	UK3	

Table 26: *Sample locations from UK, including production type, organic status, farm no. and time points where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated*

Country	Closest Town or Region	Sample name(s)	Sampling date or year	Production system (if known)
Albania	Lushnye	ALB1	2018	(Broiler) Intensive
	Durres	ALB5	2017	(Broiler) Rural farm
Belgium	Destelbergen	BEL1	2018	Intensive (Layer)
	Destelbergen	BEL2	2018	Intensive (Layer)
	Destelbergen	BEL3	2018	Intensive (Layer)
	Evergm	BEL4	2018	Intensive (Layer)
	Evergm	BEL5	2018	Intensive (Layer)
	Evergm	BEL6	2018	Intensive(Layer)
Czech Republic	Bohemia	CZH1	2018	Intensive (Layer)
Denmark	Vejle	DEN1	-	-
	Jylland	DEN2	-	-
France	Lacepede	FRA1	-	Aviary (Layer)
	Montfaucon	FRA2	-	Intensive (Layer)
	Grenade sur Garonne	FRA3	-	Intensive (Layer)
	Saint-Pons-de-Thomieres	FRA4	-	Intensive (Layer)
Germany	Hannover	GER1	2018	University flock (Laying hens)
Greece	Thessaloniki	GRC1	2018	Intensive (Layer)
	Corinth	GRC2	2018	Intensive (Layer)
	Leros	GRC3	2018	Intensive (Layer)
	Attica	GRC4	2018	Intensive (Layer)
Italy	Lecce	ITA10	20-07-18	Laying hens
	Verona	ITA13	10-05-18	Laying hens
	Brindisi	ITA14	21-04-18	Laying hens
	Verona	ITA15	18-03-18	Laying hens
	Milano	ITA16	29-03-18	Laying hens
	Pavia	ITA17	09-05-18	Breeder
Netherlands	Barneveld	NET7	2018	Intensive (Layer)
	Aalten	NET9	2018	Intensive (Layer)
Portugal	Riveria	POR1	2018	Intensive (Layer)
	Benaveute	POR2	2018	Intensive (Layer)
	Braemes	POR3	2018	Intensive (Layer)
	Souta da Carpalhosa	POR4	2018	Intensive (Layer)
	Zezeroro	POR5	2018	Intensive (Layer)
	Oliveina de Fnades	POR6	2018	Intensive (Layer)
	Ancogelo, Panti de linne	POR7	2018	Intensive (Layer)
Romania	Tatarlaua	ROM1	2018	Backyard
	Tatarlaua	ROM2	2018	Backyard
	Cuzdrioara	ROM3	2018	Backyard
	Cuzdrioara	ROM4	2018	Backyard
	Cuzdrioara	ROM5	2018	Backyard
	Floresti	ROM6	2018	Backyard
Spain	Seville	SPA4	2018	Intensive (Layer)

***Table 27**: Sample locations from Europe (outside of the UK), including the country and production type*

6.3.2 Geographical clustering

Countries were grouped into four geographical clusters based on spatial proximity and climatic factors in an arbitrary fashion, with geographic cluster ID's assigned (Table 28). This was to give a brief indication of whether patterns of diversity could be seen across Europe. The UK formed one cluster due to physical separation from mainland Europe.

Geographical cluster	Countries included
1	United Kingdom
2	Belgium, Czech Republic, Denmark, Germany and the Netherlands
3	France, Italy, Portugal and Spain
4	Albania, Greece and Romania

Table 28: Geographical clustering of countries sampled for D. gallinae *with assigned cluster number and countries included per country detailed*

6.3.3 Selection of high-quality SNPs for assay development

SNP selection was completed based on a in-house filtration system involving read depth, PHRED quality score, local BLAST alignments, SNP location and distance from other identified SNPs (Figure 37) to identify 100 high quality robust SNPs from the list generated in Chapter 5 for genotyping (referred to as 'primary SNPs'). The start goal was to identify 60 SNPs for Sequenom MassARRAY, requiring a single SNP with non-variable flanking regions. However, the Eurofins Mid-Plex genotyping assay offered a greater flexibility in number of analysed SNPs and samples as well as greater flexibility in SNP selection, hence a change in SNP filtration (Figure 37). A total of 92 primary SNPs were identified through the SNP filtration process, with a further eight identified in vaccine candidates (five) and markers of acaricide resistance (three), totalling 100 primary SNPs.

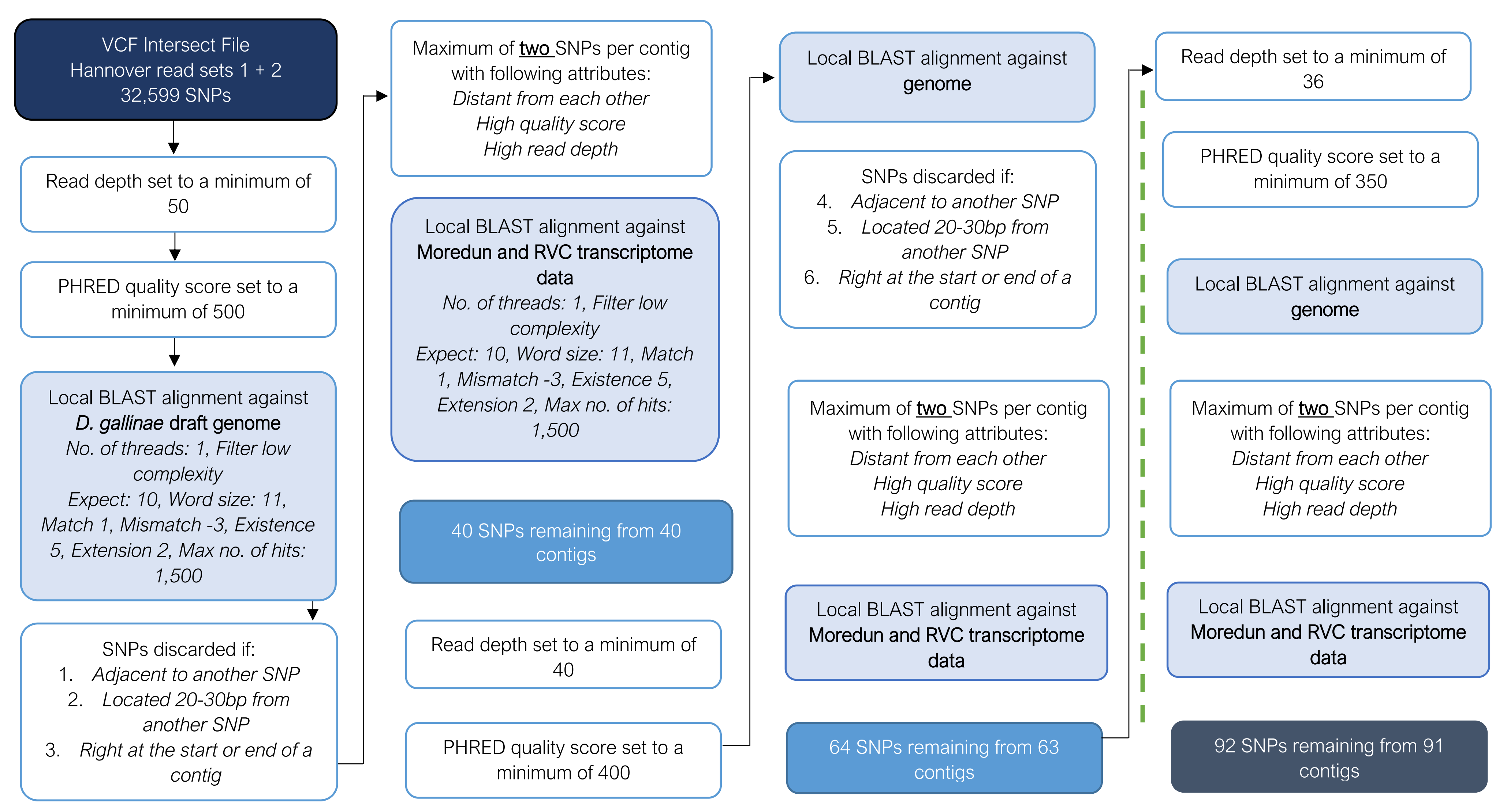


***Figure 37**: SNP filtration process of 32,599 SNPs identified in* D. gallinae *for selection of 92 target SNPs for Eurofins Mid-Plex genotyping assay. Dashed green line indicating change in plan from Sequenom specifications to Mid-plex specifications*

6.3.3.1 Incorporation of vaccine targets and acaricide resistance markers

6.3.3.1.1 Vaccine candidates

SNPs relating to four vaccine candidates (outlined in detail in Chapter eight) were selected for inclusion in the SNP panel (Figure 38). SNPs relating to tropomyosin, cathepsin-D and paramyosin were exclusively selected from GATK results. Sanger sequencing of vitellogenin fragments produced from individual *D. gallinae* DNA extracts (discussed in Chapter eight) were also used to aid SNP identification for the vitellogenin locus.

6.3.3.1.2 Acaricide resistance markers

Three SNPs identified elsewhere as genetic markers associated with acaricide resistance (outlined in detail in Chapter seven) were incorporated into the SNP panel (Figure 38). Sequences with known mutations annotated were provided by Prof. Thomas Van Leeuwen (a collaborator on mapping pyrethroid resistance (592)).

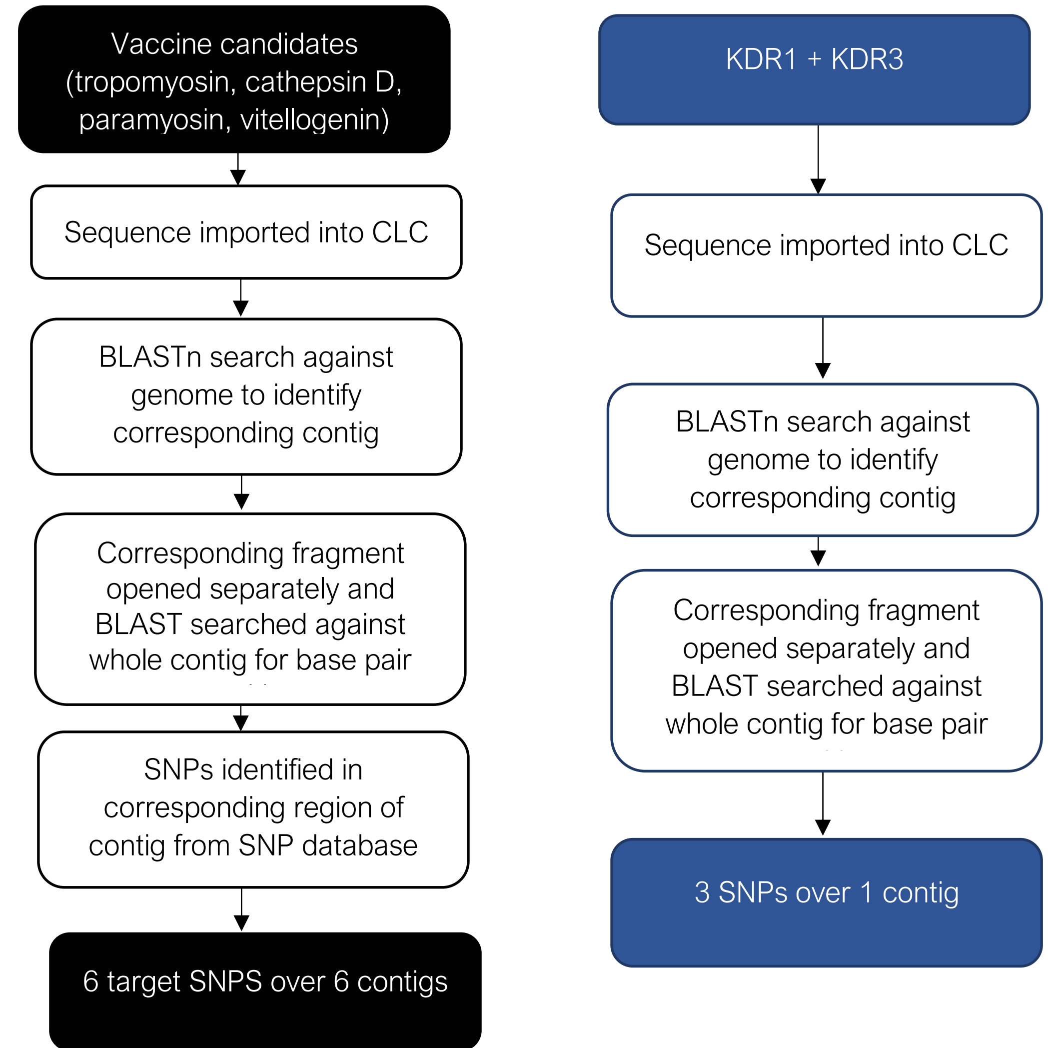


Figure 38: *Left: flowchart outlining the process of identifying SNPs in vaccine candidates for incorporation into the SNP panel Right: flowchart outlining identifying genome co-ordinates for acaricide resistance markers (KDR1 and KDR3 where KDR is kinase domain receptor. CLC: CLC Main Workbench 20.0.4 (QIAGEN)*

6.3.3.2 Accessing secondary SNP markers from the Mid-Plex sequence dataset

The Eurofins Mid-Plex genotyping assay results in amplification of 150bp upstream and downstream of the target site, creating the opportunity to incorporate additional markers (SNPs) detected within the sequenced regions flanking each primary SNP. These SNPs, referred to in this chapter as 'secondary' SNPs, were identified and used by serendipity and were not accompanied by the same quality profiles as the primary SNPs. Utilising the contig position of the 100 selected primary SNPs it was possible to identify additional secondary SNPs within 57 loci from the original data set of 32,559 (outlined in Chapter four) that fell within the corresponding 300bp that would be amplified through the Mid-Plex assay. Whilst less focus was placed on quality compared to primary SNPs, a minimum PHRED quality of score was set to 20 and read depth to 250, except for a SNP associated with vitellogenin, which was identified through Sanger sequencing, and the acaricide markers (that were selected based on previous work (592). The number of SNPs in the flanking 150bp ranged from 2 to 14, with the most common numbers being 2 or 3 additional SNPs present. An additional 57 SNPs were incorporated from the flanking regions, producing a total of 157 SNPs included in the assay.

6.3.4 Annotation of target SNPs

Contigs relating to the selected SNP markers were cross-referenced with the .gff file associated with the draft *D. gallinae* genome assembly to identify SNP locations, defining each as exonic, intronic and or associated with a known protein.

6.3.5 Validation of SNPs through PCR and Sanger sequencing

As discussed in Chapter four (Section 2.5.2), biological validation was important for confirmation of computationally identified SNPs to ensure SNPs were genuine. In order to achieve this, a subset of ten SNPs were selected, spanning a range of quality score and read depth values, for PCR amplification and Sanger sequencing.

6.3.5.1 Primer design

A total of twenty primers were designed to cover all ten SNPs selected for validation (Table 29). Primers were designed as described in General Methodology 2.6 Primer design and use.

6.3.5.2 PCR and agarose gel electrophoresis

PCR and agarose gel electrophoresis were carried out following the protocols outlined in General Methodology 2.7 and 2.9, respectively.

6.3.5.3 PCR purification and Sanger sequencing

PCR purification and Sanger sequencing were carried out following the protocols outlined in General Methodology 2.10 and 2.11, respectively.

Primer name	Primer sequence (5'-3')	Length	TM (°C)	GC%	Secondary structure	Primer dimer
3867F	GCAAATCGAGTTTAGTGACCCA	22	65.0	45.4	None	None
3867R	AGCATCTTCATCAGCCATCAAG	22	65.4	45.4	Very Weak	None
5958F	GTTGTAAGTCGCGTGGAATCCG	22	69.5	54.5	None	None
5958R	ACGGCAAGCCATTTCTATCGTA	22	66.4	45.4	Weak	None
5193F	GCGAACAAATCGTATGAGCAAC	22	65.4	45.4	Weak	None
5193R	CTTTCTCCACGTCAACAGCGCTT	23	70.5	52.1	None	None
2130F	GCTCATTTCATAGGGGATGATG	22	64.3	45.4	Weak	None
2130R	TGCTTAACATTCACACCTTTGT	22	61.0	36.3	Very Weak	None
6526F	ACTAATAGCGCTGATTCTCTGGAT	24	63.1	41.6	None	None
6526R	GTAACAACTCGTAATCTCGCGA	22	63.0	45.4	Weak	None
5655F	GACAGTGTATGTTTGCGCATGT	22	64.8	45.4	Very Weak	None
5655R	GATCACTTTCGAAGGGAGGGAG	22	67.1	54.5	Very Weak	None
2114F	TCAAAAGGGCAGAGATGTTCTT	22	63.8	40.9	Very Weak	None
2114R	TCTTTCAAACCGAACGTCATCC	22	67.1	45.4	Very Weak	None
2307R	AAAGTTTGGTGATGTGGGCTAA	22	64.3	40.9	None	None
2307R	CATATACGCTATGATCGGCACT	22	62.7	45.4	Weak	None
1595F	CCGAACTTCAAACAGGCCGTAA	22	68.9	50	None	None
1595R	CAGGTCATACACACGCCAACAA	22	67.9	50	None	None
6416R	ACCTTGTAGAATTTGAGCACGG	22	64.1	45.4	None	None
6416R	TCTAAGTGTATGCGTTCGTTCA	22	62.2	40.9	None	None

Table 29*: Primer sequences, length, melting temperature, GC%, presence of secondary structure or primer dimer in primers used to validate ten SNP markers within D. gallinae genomes.*

6.3.6 Eurofins Mid-Plex Genotyping method

Genotyping was conducted by Eurofins Genomics using their Mid-Plex genotyping method and analysed using Mix-Plex analysis software (Eurofins Genomics Europe Sequencing GmbH, Konstanz, Germany).

6.3.6.1 Assay design

To develop the primers required for the Mid-Plex genotyping a 300bp sequence fragment with corresponding SNP ID and annotated SNPs were provided to Eurofins Genomics to allow for primer design. Primers were designed by Eurofins and remain their proprietary property. Multiple marker targets were identified and ten SNPs with the highest priority were labelled (Mite_0881, Mite_2307, Mite_5781 and the seven target sequences relating to vaccine candidates (vitellogenin, tropomyosin, paramyosin and cathepsin-D)) and two acaricide resistance related markers (KDR1 and KDR3).

6.3.6.2 Sample preparation

Samples were prepared in line with the sample preparation guidelines provided by Eurofins Genomics. Briefly, a total of 20µl of extracted DNA was provided per sample, eluted in RNA-free ultra-purified water. All DNA samples were treated with RNase, to minimise RNA contamination, as outlined in General Methodology 2.4. DNA concentration was measured using a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) with a range of 1.961 to 55.208 ng/µl (average 11.698 ng/µl), as outlined in General Methodology 2.5.1 or using a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts) with a range of 1.000-141.592 ng/µl (average 22.261 ng/µl) as outlined in General Methodology 2.5.2. Samples were loaded into skirted Eppendorf twin.tec® PCR plates (Eppendorf, Hamburg, Germany), leaving G12 and H12 empty (to be used as controls) and sealed with Eppendorf PCR cap strips (eight strips, domed) (Eppendorf, Hamburg, Germany). Plates were sent on dry-ice via 24 hour courier service direct to Eurofins Genomics (Eurofins Genomics Europe Sequencing GmbH, Konstanz, Germany) for analysis. Corresponding sample spreadsheets were provided outlining plate barcode, sample number, sample position, well position, sample name, volume and concentration.

6.3.6.3 Assay optimisation

During assay optimisation, Eurofins genomics adjusted their genotyping methodology to factor in low DNA concentration across samples. Three proprietary primer mixes were analysed to provide output relating to coverage class achieved (Figures 39-40). Based on the results, primer

mix version two was selected for all samples due to achieving the highest yield of coverage classes over 100x.

6.3.7 Data processing

Three criteria were applied to format the raw SNP genotype data on receipt from Eurofins. Firstly, the minimum read cut off was set to 3 for a genotype to be called and included. Secondly, only samples with a SNP call rate of 70% or higher were included. Finally, only SNPs called from 90% of samples or higher were retained for use.

6.3.7.1 Conversion of heterozygous allele calls to dominant allele

Heterozygous calls (assigned the identifier HET by Eurofins) were converted to the dominant allele using raw allele read depths generated during sequencing. To achieve this, the allelic fraction was calculated based on ALT/(REF+ALT), where REF = reference allele read depth and ALT = alternative allele read depth. A ratio of 0.6 or higher was used to infer alternative alleles and a ratio lower than 0.6 used to infer reference alleles.

6.3.8 Population genetic analyses

6.3.8.1 Nucleotide analysis

The number of SNP haplotypes were calculated for the entire data, by geographical clustering and by production system across Europe, UK production systems and UK organic status using DnaSP version 6.12.03 (593).

6.3.8.2 Linkage disequilibrium analysis

Linkage equilibrium is characterised by statistical independence of alleles between loci (594). LIAN 3.0 was used to analyse haplotype linkage disequilibrium (LD) across *D. gallinae* populations (594). LIAN, short for LInkage ANalysis, tests for independent assortment by first computing the number of loci at which each pair of haplotypes differ. From the distribution of mismatch values a variance, V_D, is calculated, which is compared to the variance expected for linkage equilibrium, V_e. The null hypothesis $H_0 : V_D = V_e$ was tested by Monte Carlo simulation and by a new parametric test (595).

Further linkage disequilibrium analysis was conducted using DnaSP (Version 6) (593) to identify the number of SNP markers which demonstrated with significant LD. The analysis was performed using a pairwise comparison for all polymorphic sites in the data, with statistical significance computed using the chi-square test with Bonferroni correction for multiple analyses applied.

6.3.8.3 *Network analysis*

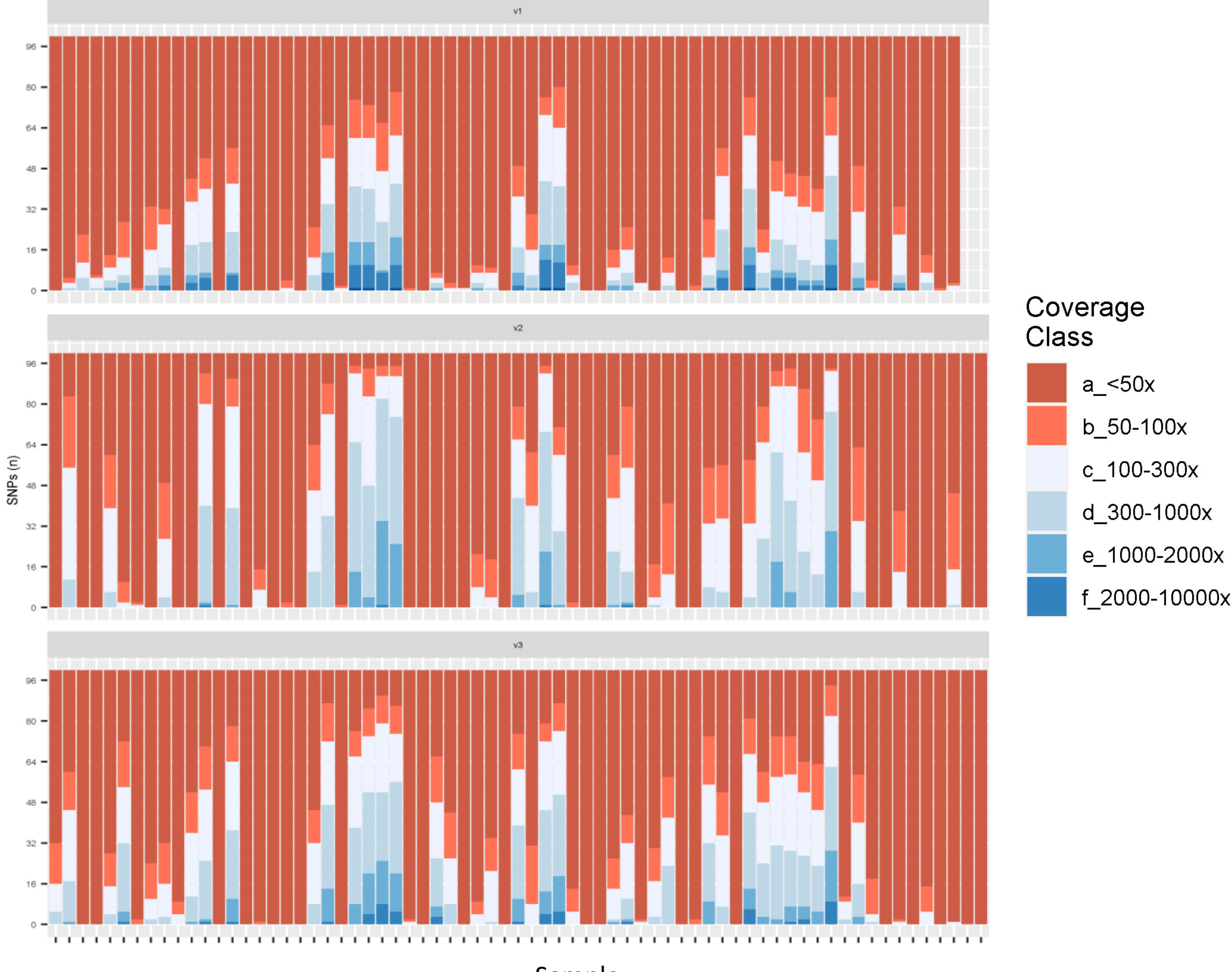
Network 5.0.0.3 (www.fluxus-engineering.com) was used to construct Median-Joining (MJ) trees (359). Mites with identical SNP sequences were designated as one haplotype. Nodes in each network were colour coded to represent whole countries to provide a visual indication of the relationship of haplotypes within and between countries.

6.4 RESULTS

A total of 108 pooled *D. gallinae* samples and 157 SNPs were analysed using the Eurofins Mid-Plex genotyping method, with SNP identities called initially using Eurofins Genomic Mid-Plex analysis software. Outputs included (i) read depth for the reference (REF) or alternative (ALT) SNP type at each locus, supplemented by (ii) a genotype call (REF, ALT or heterozygous (HET)).

6.4.1 Assay optimisation

Results of three different proprietary primer mixes used on 100 primary SNP loci demonstrated that use of primer mix version two produced the best results for coverage across the samples tested (Figures 39, 40).



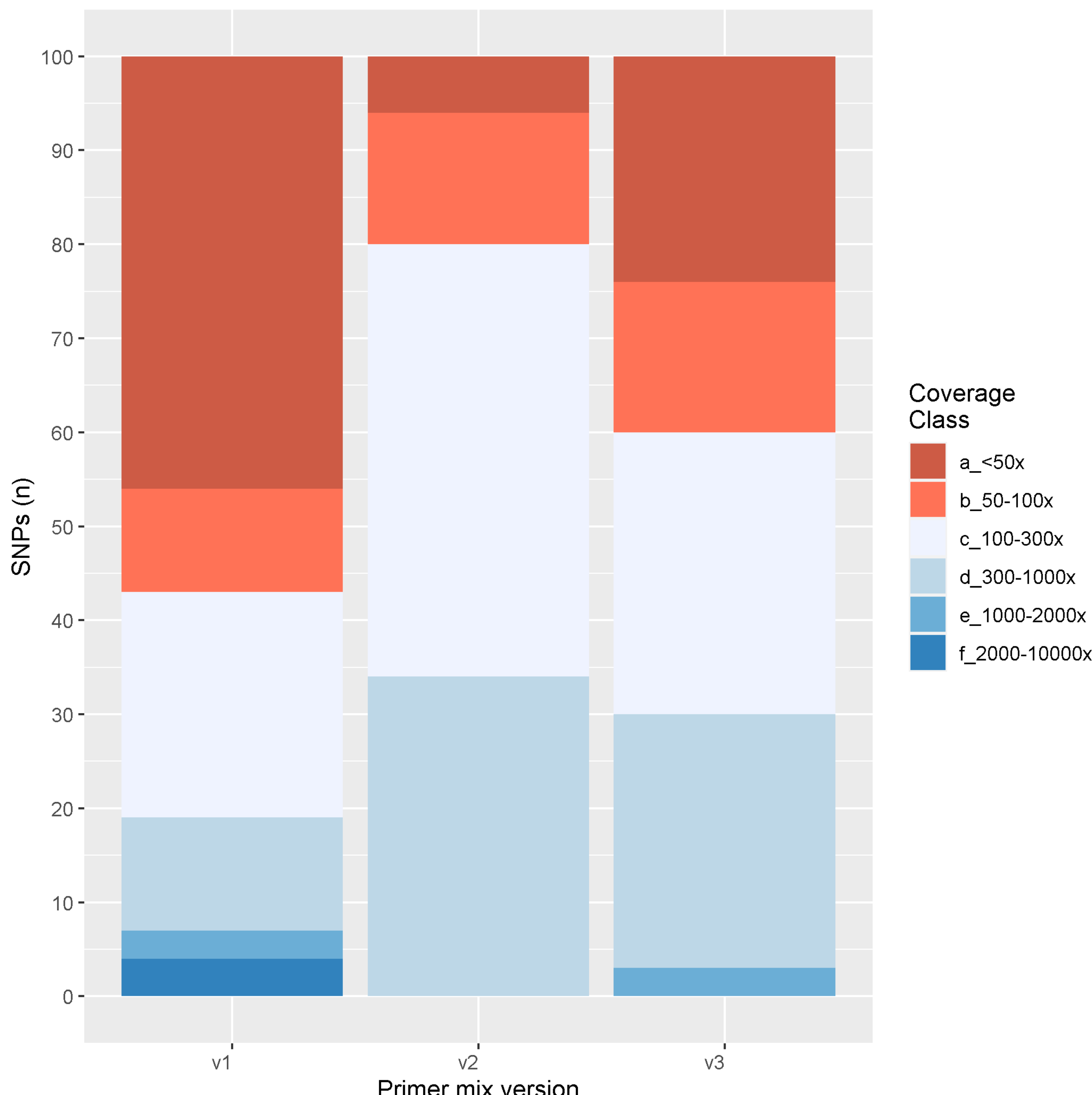
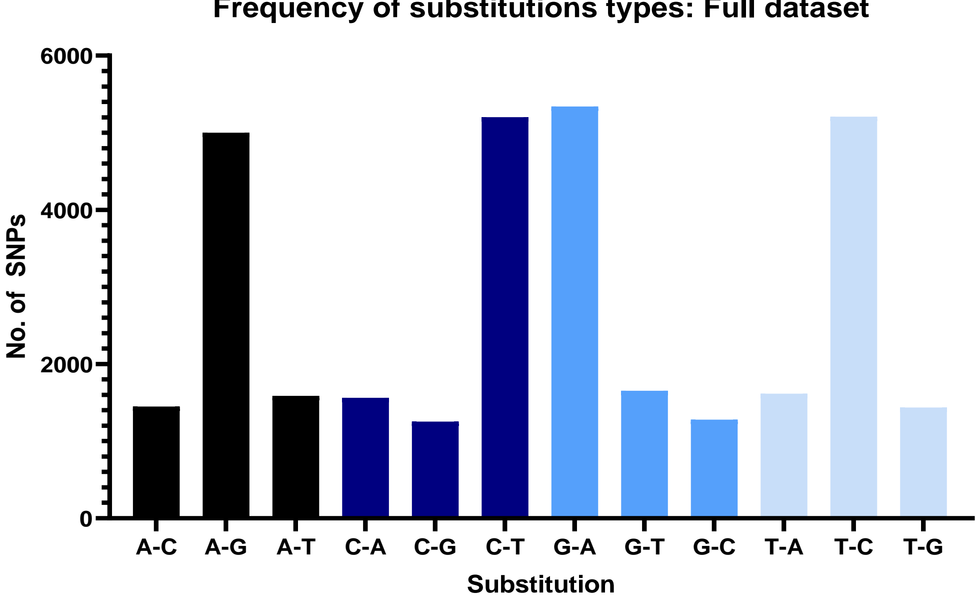


***Figure 39**: Coverage classes achieved by proprietary primer mixes for individual samples tested on pooled* D. gallinae *samples for 100 primary SNPs by Eurofins Genomics during assay optimisation. Sample names have been removed for confidentiality, with one sample present per column, separation indicated by black lines (I).*

***Figure 40:** Total coverage classes achieved by proprietary primer mixes tested on pooled* D. gallinae *samples for 100 primary SNPs by Eurofins Genomics during assay optimisation*

6.4.2 Marker characteristics for the primary SNPs selected for Mid-Plex analysis

Quality scores ranged from 28.07 to 2552.41, with an average score of 698.49. Read depth scores ranged from 2 to 127, with an average score of 50. Comparison of substitution frequency between the full dataset of 32,599 SNPs and the targeted SNPs revealed a similar profile (Figure 41).



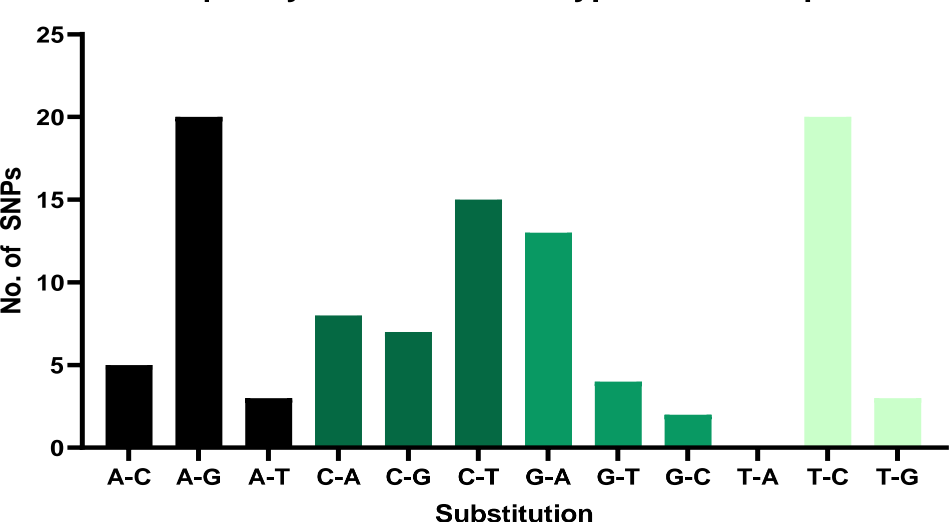


Figure 41: Frequency of substitution changes. Top: Frequency for entire dataset Bottom: Frequency of target 100 SNPs

6.4.2.1 Intron/Exon Annotation

Out of 100 SNPs chosen, 25 were not annotated in the .gff file, 61 related to exons and 14 to introns (based on being found in a region coding for a gene but not associated with any of the available exons) (Figure 42).

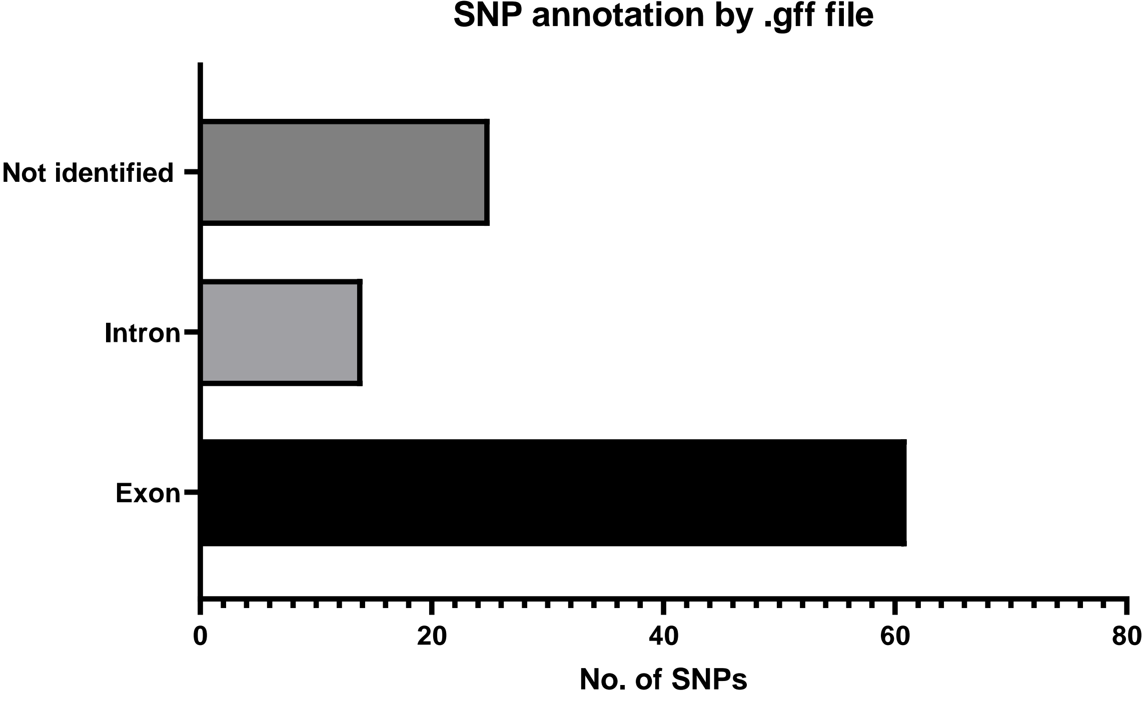


Figure 42: *Annotation associated with the primary SNPs identified in* D. gallinae *and selected for SNP genotyping including exon, intron or not identified*

Two SNPs were predicted to be located in the five prime UTR region of their associated gene and nine were associated with the three prime UTR. 16 SNPs were in genes with homologues identified in other organisms, according to the algorithms utilised to produce the .gff file, with the closest related species being *Ixodes scapularis* (Table 30). From this 16, four SNPs were located in the intronic region of their associated gene and the remaining twelve located in exonic regions. One SNP was predicted to be found in the intron of one gene on the forward strand and in an exon of a different gene in the reverse strand. All SNPs relating to the vaccine candidates were identified to be in exons but no corresponding homologues were annotated in the .gff file. Investigation into the 25 SNPs which failed to link to the .gff file for annotation revealed no notable difference in quality score or read depth between annotated and non-annotated SNPs, suggesting that the pipeline produced robust SNPs but the annotation was missing in the .gff file at this current time (Figure 43).

#CHROM	POS	REF	ALT	.gff file					
				SOURCE	FEATURE	START	END	STRAND	COMMENT
Mite_0333	13576	C	T	Maker	Exon	13509	13665	+	Similar to PSMB8: Proteasome subunit beta type-8 (*Canis familiaris*)
Mite_0534	457483	A	T	Maker	Intron	442784	457947	-	Similar to PXN: Paxillin (*Pongo abelii*)
Mite_1387	470677	T	C	Maker	Intron	459680	471350	+	Similar to CWC22: Pre-mRNA-splicing factor CWC22 homolog (*Gallus gallus*)
Mite_1595	1215622	T	C	Maker	Exon	1210802	1215793	+	Similar to rpl-7: 60S ribosomal protein L7 (*Caenorhabditis elegans*)
Mite_1622	477706	A	G	Maker	Exon	475046	483902	+	Similar to Vps28: Vacuolar protein sorting-associated protein 28 homolog (*Rattus norvegicus*)
Mite_1930	178395	T	G	Maker	Intron	172795	182568	-	Similar to Cuticle protein 10.9 (*Ixodes ricinus*)
Mite_2189	2631219	C	T	Maker	Exon	2626210	2631779	-	Similar to Gstm3: Glutathione S-transferase Yb-3 (*Rattus norvegicus*)
Mite_2754	124615	C	T	Maker	Exon	118982	124740	-	Similar to ATPsyn-d: ATP synthase subunit d%2C mitochondrial (*Drosophila melanogaster*)
Mite_3240	247753	G	C	Maker	Exon	246938	264262	-	Similar to RpS8: 40S ribosomal protein S8 (*Spodoptera frugiperda*)
Mite_3487	36926	C	G	Maker	Intron	30362	40170	+	Similar to hd: Protein downstream neighbor of son homolog (*Drosophila melanogaster*)
					Exon	35090	44663	-	Similar to GSTO1: Glutathione S-transferase omega-1 (*Homo sapiens*)
Mite_4723	14813	C	A	Maker	Exon	11129	14976	-	Similar to RNA-binding protein pno1 (*Ixodes scapularis*)
Mite_5655	184497	G	A	Maker	Exon	184292	187187	-	Similar to RBG4: Glycine-rich RNA-binding protein 4%2C mitochondrial (*Arabidopsis thaliana*)
Mite_5794	373391	T	C	Maker	Exon	372999	385925	-	Similar to prps2: Ribose-phosphate pyrophosphokinase 2 (*Xenopus laevis*)
Mite_5866	2163501	A	G	Maker	Exon	2163228	2164233	-	Similar to RpL19: 60S ribosomal protein L19 (*Drosophila melanogaster*)
Mite_2341	216525	C	T	Maker	Intron	132865	226245	+	Similar to RPL39P5: Putative 60S ribosomal protein L39-like 5 (*Homo sapiens*)
Mite_2272	53506	A	G	Maker	Exon	48211	55968	-	Similar to CLTC: Clathrin heavy chain 1 (*Homo sapiens*)

Table 30: *Information regarding associated proteins identified in the .gff file, including chromosome number, position relative to contig, reference and alternative alleles, source, feature type (intron or exon), start and end of contig, score, strand (positive or negative), frame and comment on similarity to other known organisms*

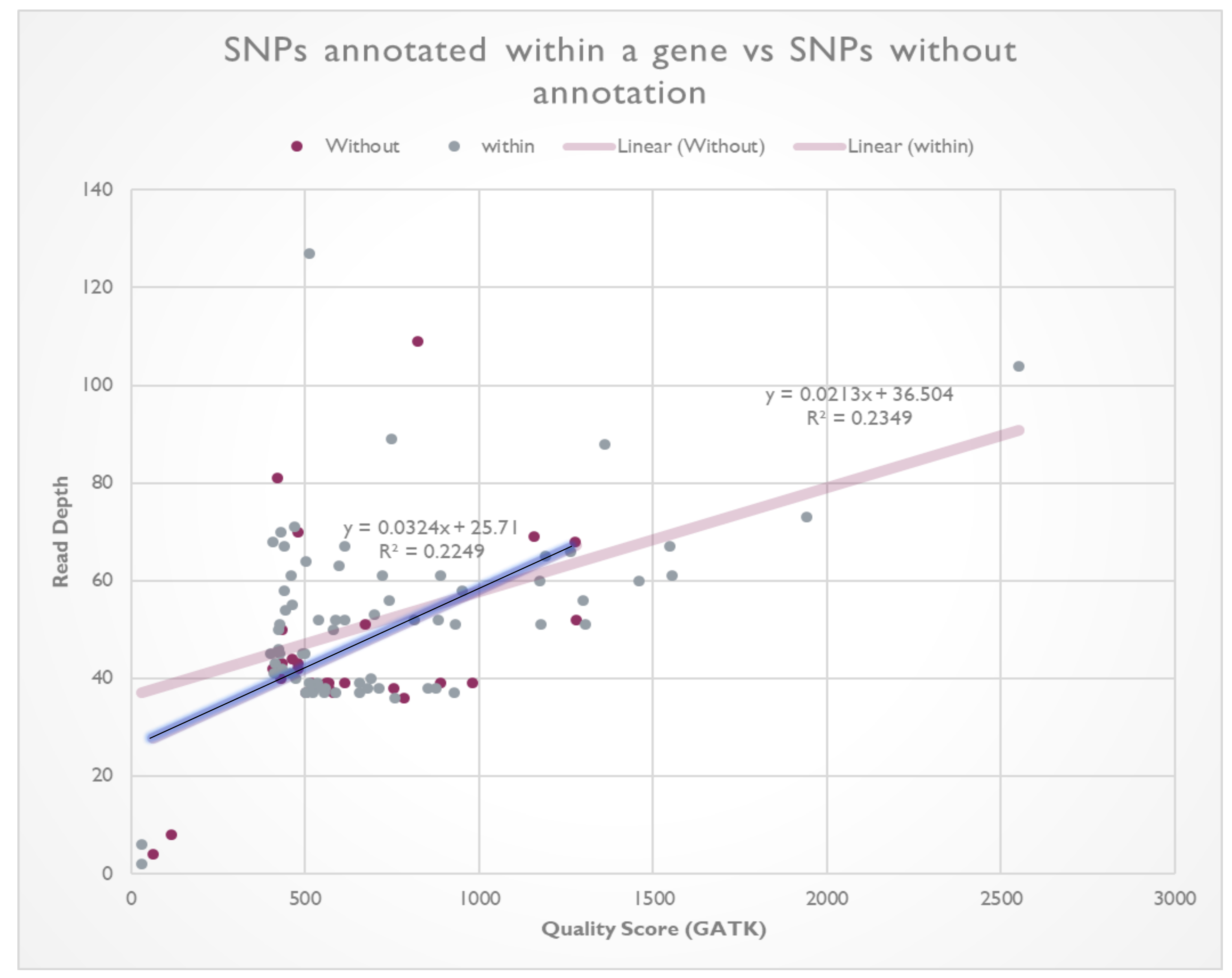


Figure 43: *Quality score and read depth scores for SNPs with corresponding .gff annotation vs. SNPs with no annotation found. Red dots representing SNPs with no associated annotation and grey representing SNPs associated with an annotation in the .gff file*

6.4.3 Marker characteristics for the secondary SNPs selected for Mid-Plex analysis

A total of 57 secondary SNPs were included in Mid-plex genotyping analysis, with quality scores and read depth data for 47 available. Ten secondary SNPs were identified through Sanger sequencing of vitellogenin fragments (discussed in greater detail in Chapter 6). As they were not identified in the GATK pipeline, read depth and quality scores were not available and as such are not reflected in the averages. Quality scores ranged from 112.2 to 1257.65, with an average score of 646.95. Read depth scores ranged from 8 to 80, with an average score of 45. Compared to the primary SNPs, a reduction in average quality score of 51.54 and read depth of 5, indicating that quality of additional markers remained high.

6.4.4 Biological validation: PCR results

Results from amplification of ten SNP markers in pooled populations provided biological validation for seven of the computationally identified primary SNPs as being present and variable in *D. gallinae* populations (UK6, UK11, UK9, ROM2, BEL1, POR2) (Table 31). Two primer pairs produced a visible band on the agarose gel, but sequencing failed to yield results, while only alternative alleles were observed for the tenth SNP (#2307; Table 31).

Primer pair	Successful		Reference and Alternative alleles present
	Amplification	Sequencing	
3867	Yes	No	-
5958	Yes	No	-
5193	Yes	Yes	Yes
2130	Yes	Yes	Yes
6526	Yes	Yes	Yes
5655	Yes	Yes	Yes
2114	Yes	Yes	Yes
2307	Yes	Yes	No (only alternative)
1595	Yes	Yes	Yes
6416	Yes	Yes	Yes

Table 31: *Summary of PCR results for biological validation of ten SNP markers from pooled* D. gallinae *samples from four farms, including amplification success and presence of reference and alternative alleles*

6.4.5 Summary statistics

A total of 157 SNPs were analysed for 108 pooled *D. gallinae* samples achieving a 76.6% call rate (Table 32). From the 108 samples, a call rate of >90% was achieved for over half of samples (65.7%), whilst <10% of SNPs were successfully called for 14 samples (Table 32). From the 157 SNP markers, almost half achieved a call rate >80% (47.2%) and four samples were considered a failure with less than <10% call rate (Table 32). In terms of heterozygosity, 143 markers were represented by at least one heterozygote.

Number of analysed SNPs:	157
Number of analysed SAMPLES:	108
Overall calling rate:	76.6%
Samples	
Number of SAMPLES with >90% call rate:	71 (65.7%)
Number of SAMPLES between ≥50% and >90% call rate	17 (15.7%)
Number of SAMPLES between >10% and <50% call rate	6 (5.5%)
Number of SAMPLES with <10% call rate	14 (12.9%)
SNPs	
Number of SNPs with >80% call rate	74 (47.2%)
Number of SNPs between ≥50% and <80% call rate	79 (50.3%)
Number of SNPs between >10% and <50% call rate	4 (2.5%)
Number of SNPs with <10% call rate	0
Total no. of genotypes called	
Reference	8884
Alternative	4319

Table 32: *Summary genotyping results of 157 SNPs for 108 pooled D. gallinae samples from across Europe and UK with percentage call rates achieved for samples and SNPs*

6.4.5.1 Variation in SNP markers

After quality control to the criteria outlined in 2.3.7, a total of 145 SNPs and 75 samples were included in the analyses. Overall, 132 markers were polymorphic, 12 markers only called reference alleles and one marker only called alternative alleles for a total of 75 pooled *D. gallinae* samples (Table 33). Quartile ranges demonstrate 77 markers with 25-75% frequency for reference and alternative alleles, indicating these as informative markers in *D. gallinae* (Table 33).

		No. of SNPs	Associated Marker
Non-variant sites			
(All) Reference		12	Mite 4456_1, Mite 4456_2, Mite 5400_6, Mite 1635_2, Mite 1635_3, Mite 3487_2, Mite 6481_1, Mite_1011_1, Mite 2196_1, Mite 2754_1, Mite 4872_1, Mite 5781_1
(All) Alternative		1	Mite 1414
Quartile ranges			
Reference	<25%	5	
	25-50%	38	
	50-75%	39	
	75-100%	63	
Alternative	<25%	63	
	25-50%	40	
	50-75%	36	
	75-100%	6	

Table 33: *Summary of 145 SNPs for 75 pooled* D. gallinae *samples retained for analysis after quality control, including, alternative or heterozygotes called and the quartile ranges for reference and alternative NA alleles. Discrepancies in the quartile ranges due to uncalled markers (i.e. N's)*

6.4.6 Nucleotide analysis

Mean genetic diversity, calculated using LIAN (594), for all samples combined was 0.3252 (Table 34). Based on geographical cluster, cluster two had the lowest diversity (0.2700) whilst cluster three had the highest (0.3147). Comparing production systems across Europe, free-range farms demonstrated marginally higher diversity without statistical significance. Across all populations, no shared haplotypes were observed with a unique haplotype recorded for every sample (Table 34).

6.4.7 Linkage disequilibrium

Linkage analysis utilising LIAN revealed significant linkage disequilibrium in the full dataset, all geographic clusters analysed, and all production systems except for UK intensive farms (Table 34). Linkage disequilibrium (LD) analysis completed using DnaSP included 4005 pairwise comparisons and revealed a total of 801 significant pairwise comparisons by chi-square test, 83 or which remained significant using the Bonferroni procedure. The coefficient of LD (D) was normalised in DnaSP and presented as D' based upon the theoretical maximum difference between the observed and expected haplotype frequencies. D' is a standardisation method which is calculated by the comparison of D compared to its maximum through the following equation $D' = D/Dmax$. D' was positive at all statistically significant sites, with and without the Bonferroni correction (Table 35), indicating that these markers occurred together in the same haplotype more than expected. Across all significant sites, 82 markers showed significant LD but only 30 markers showed significance after Bonferroni correction (Table 35).

Dataset analysed	No. of samples	No. of haplotypes	*H*	V_D	V_e	I_A^S	Var(V_D)	*P*	*L*
All	75	75	0.3252 +/- 0.0149	127.2674	27.3605	0.0252	2.4292	<0.01	30.5457
Production type*									
Free-range/Rural farm/Private flock	32	32	0.3164 +/- 0.0167	147.1455	25.6479	0.0327	4.6002	<0.01	29.3277
Intensive systems	39	39	0.3105 +/- 0.0145	109.2708	26.7990	0.0212	7.1231	<0.01	32.1939
Geographic cluster									
1 (UK)	28	28	0.2978 +/- 0.0164	131.1537	24.8623	0.0295	7.9146	<0.01	30.5562
2	13	13	0.2700 +/- 0.0162	113.7663	24.2043	0.0242	42.7805	<0.01	36.9515
3	22	22	0.3147 +/- 0.0151	92.2186	26.6902	0.0169	17.4109	<0.01	34.8881
4	12	12	0.2992 +/- 0.0185	267.1434	23.3354	0.0721	19.6625	<0.01	32.5587
Production type (UK)									
Intensive	3	3	0.2851 +/- 0.0275	44.3333	13.7778	0.0154	400.8989	0.12	65.3333
Free-range	25	25	0.2972 +/- 0.0166	130.7141	24.6512	0.0297	9.3942	<0.01	30.0721
Organic status (UK) **									
Organic	10	10	0.2753 +/- 0.0184	151.7091	21.9852	0.0407	29.0885	<0.01	32.7091
Non-organic	15	15	0.3026 +/- 0.0169	151.3106	24.7913	0.0352	23.0132	<0.01	35.7308

Table 34: *Linkage analysis of 145 SNP markers for 75 pooled* D. gallinae *samples using LIAN (594). Results shown for each dataset analysed including the full dataset, production systems across Europe, four geographic clusters, UK production system and UK organic status. Number of samples and number of haplotypes included for each dataset provided *Four farms were removed from production type analysis due to information regarding production system utilised not avaliable. ** Total of 25 farms included in organic status analysis, as intensive farms due to inability to be organic. H : mean genetic diversity, V_D: observed mismatch variance, V_e: expected mismatch variance, I_S^A: Standardized index of association, P: calculated significance, L: Simulated 5% critical value*

Significance	No. of markers	Average D’
All significant sites: P<0.001 – P<0.05	82	0.04323
P<0.001, B	30	0.50378

Table 35: *Linkage disequilibrium across 140 markers genotyped for 75 D. gallinae populations calculated on DNAsp, with all significant sites and Bonferroni corrected sites average D' and number of associated markers*

6.4.8 Network analysis of all genotyped samples

A total of 75 haplotypes were identified through network analysis, corresponding to the 75 pooled samples analysed, with no shared haplotypes observed. Clustering of samples originating from the same country was observed for a few countries, including Belgium and Romania (Figure 44). Conversely, other countries, including the UK, are distributed across the network map, sharing phylogenetic relatedness with multiple countries.

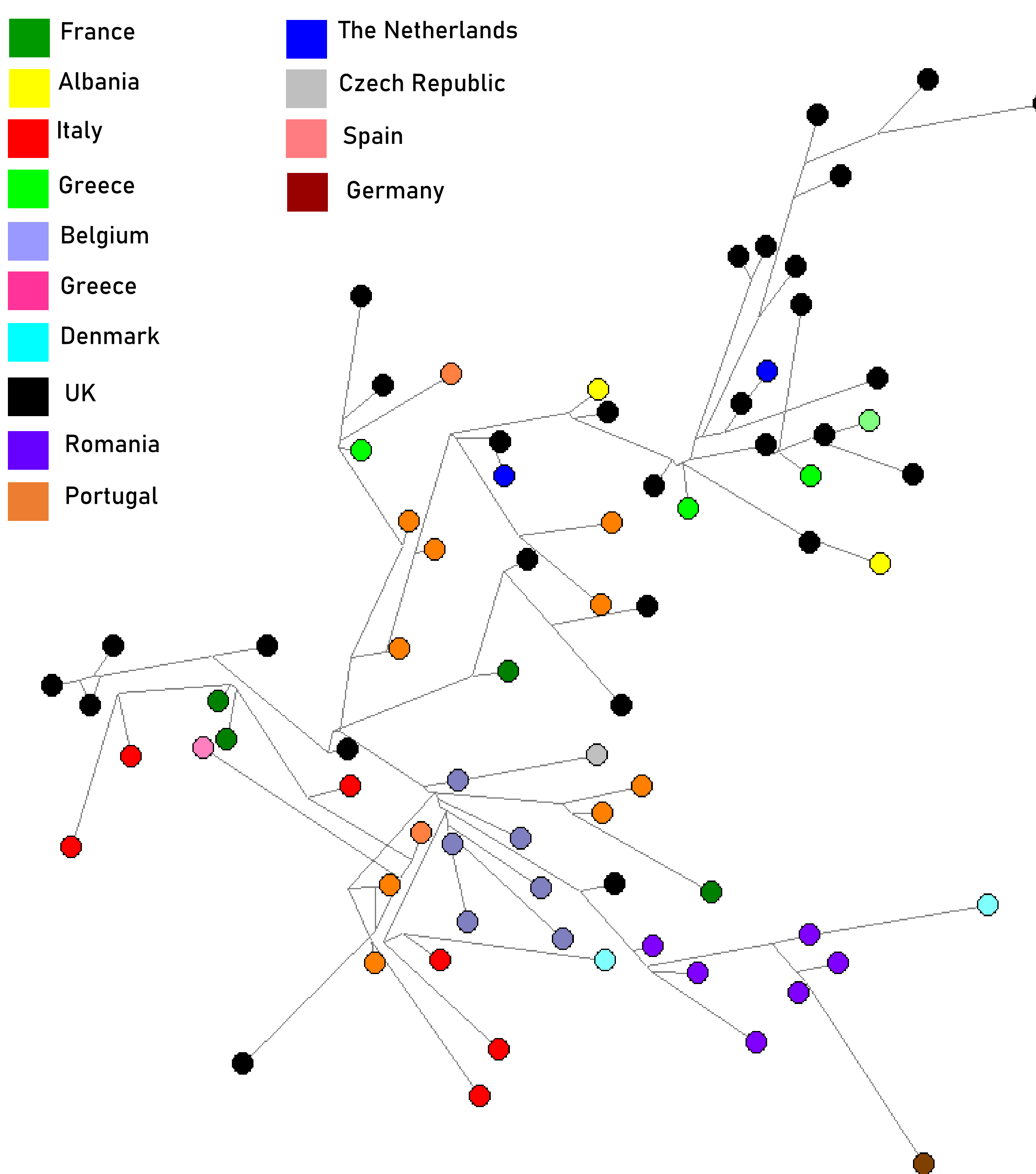


Figure 44: *Network analysis of pooled* D. gallinae *samples from the UK and the rest of Europe genotyped by NGS multiplex sequencing (237). Countries are colour coded in the key. Median-joining tree analysis was completed on Network 10.0.0.0. A total of 145 positions were used in this analysis, encoding 75 nucleotide sequences*

6.4.9 Network analysis of UK farms

6.4.9.1 Comparison between individual UK farms – spatial variation

Network analysis of 29 UK farms demonstrated 29 haplotypes divided into two haplogroups, A and B, with 17 haplotypes in group A from 12 UK farms and 12 haplotypes in group B from seven UK farms. UK7 was the only farm to have haplotypes in each haplogroup, from different time points sampled. Haplogroup B was predominately made up of haplotypes from UK6 collected over three years, accounting for 50% of the haplotypes (Figure 45).

6.4.9.1.1 Comparison between UK farms over time - temporal variation

Haplotypes recorded from samples collected from farm UK6 on four occasions at 14, 15 and 20 months after the initial collection clustered in haplogroup B, with distinct haplotypes seen for each time point sampled (Figure 45). Samples 1-4 (spanning 2017-2020) were all collected from Barn 1, with Barns 2 and 3 sampled only in 2020 at the same time as sample 4 (labelled samples 5 and 6). Distinct haplotypes are observed across the three barns with a close relationship demonstrated between time points 2-4 on the network map, whilst populations from Barn 2 and Barn 3 were more closely related to UK7 and UK18, respectively. Sampling from 2017 revealed a closer phylogenetic relationship to UK13 than to other haplotypes from UK6. Samples from UK7 on different occasions for temporal analysis had distinct haplotypes and were found in both haplogroups, whilst samples from UK11 clustered closely together in haplogroup A.

6.4.9.2 UK nations

Network analysis based on nations within the UK revealed phylogenetic relationships between farms across the UK, with haplotypes from England, Northern Ireland, Wales and Scotland found in haplogroup B. Haplotypes from Northern Ireland, Scotland and England were found in haplogroup A, with clustering of farms from Northern Ireland observed (Figure 45).

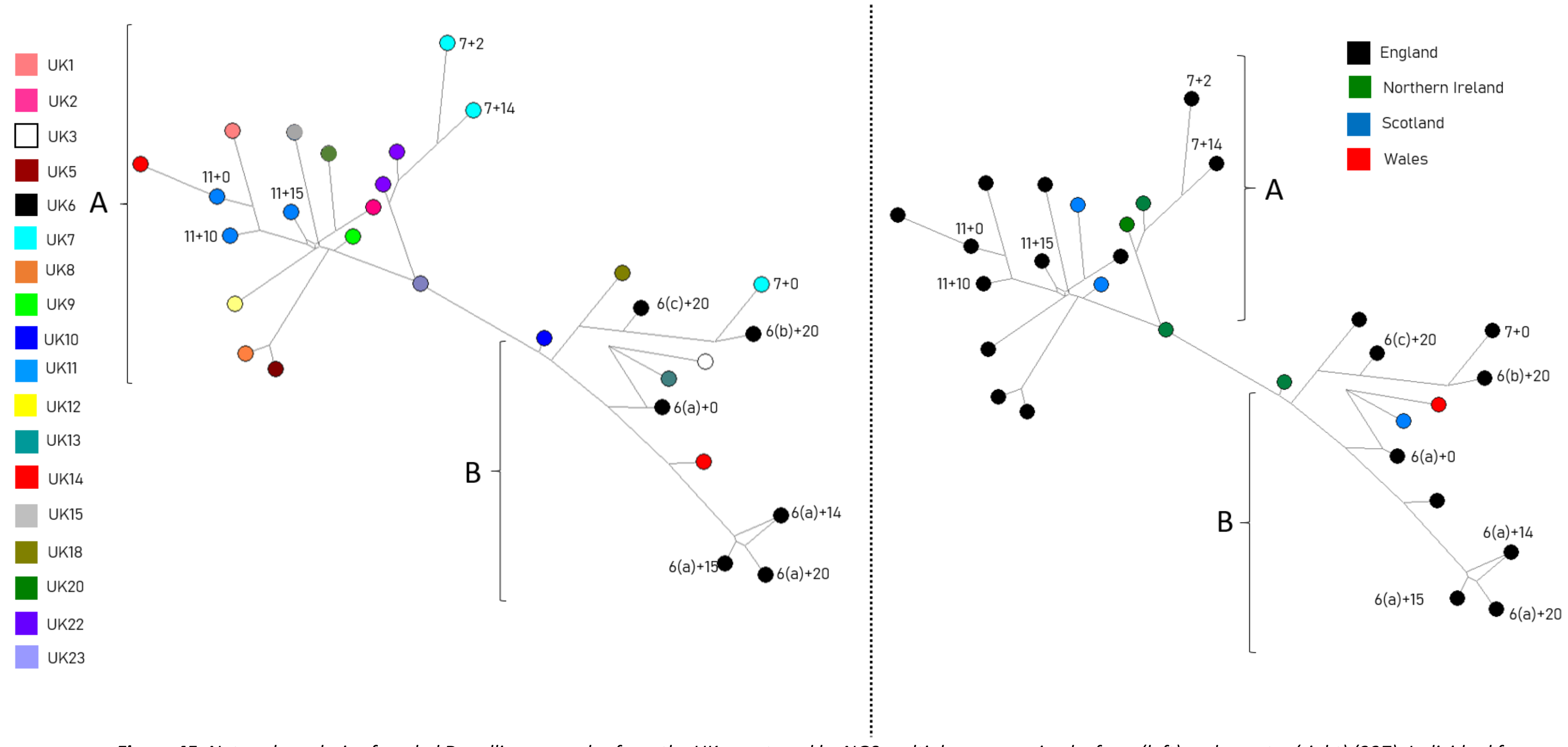


Figure 45: *Network analysis of pooled D. gallinae samples from the UK genotyped by NGS multiplex sequencing by farm (left) and country (right) (237). Individual farms and countries are colour coded in the keys provided. Farms UK6, UK7 and UK 11were sampled on multiple occasions, permitting analysis of temporal variation, where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated. Multiple barns sampled for UK6 are indicated by letters (a) barn 1 (b) barn 2 and (c) barn 3. Median-joining tree analysis was completed on Network 10.0.0.0. A total of 145 positions were used in this analysis, encoding 28 nucleotide sequences*

6.4.9.3 Network analysis of UK production systems

Analysis of production systems in the UK demonstrated close phylogenetic relationships between all production systems and no clear differentiation between organic or non-organic free-range farms (Figure 46). Haplotypes from free-range (organic and non-organic) and intensive farms were demonstrated in both haplogroups, with a higher number of haplotypes from intensive and organic farms in haplogroup A.

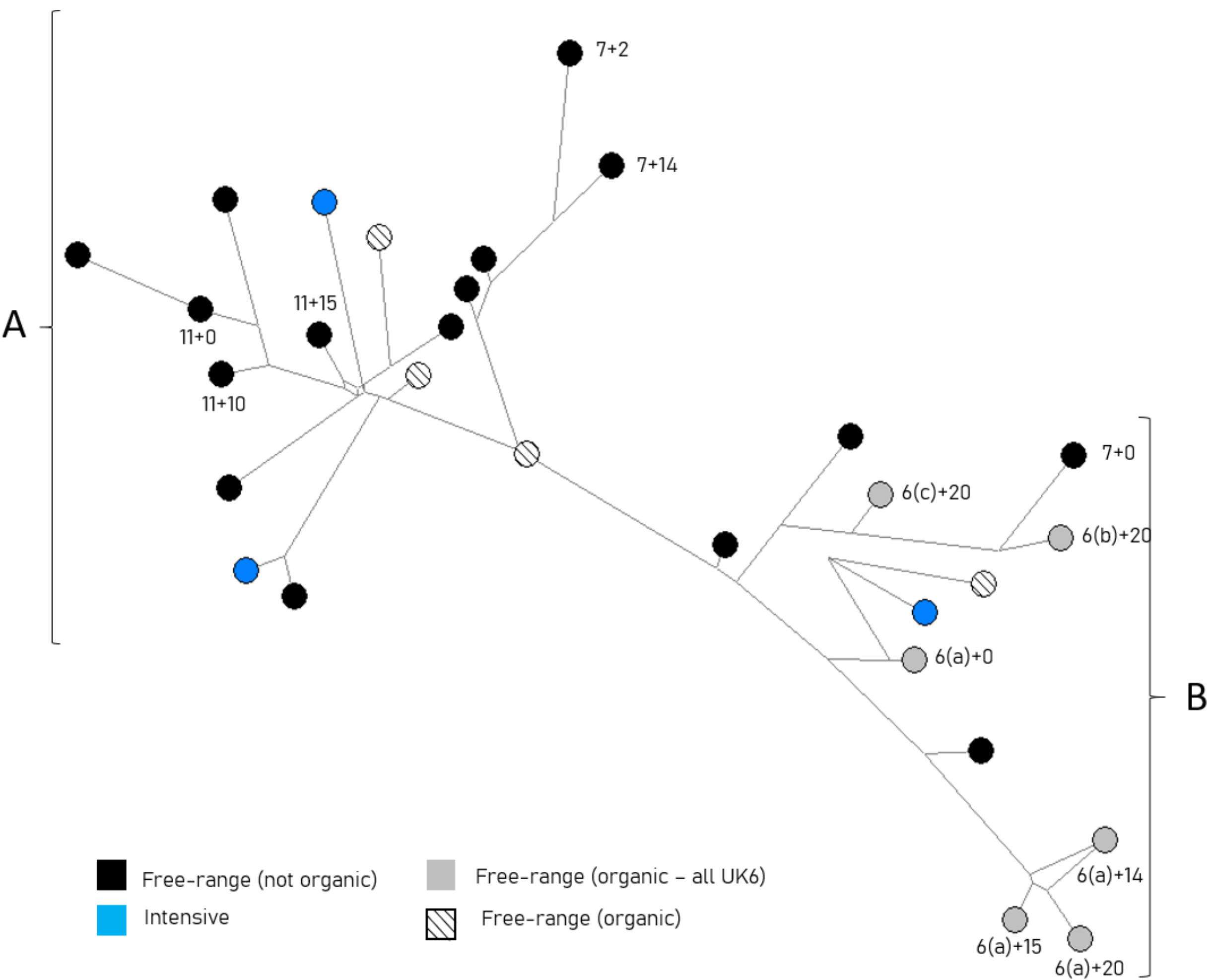


***Figure 46**: Network analysis of pooled D. gallinae samples from the UK genotyped by NGS multiplex sequencing BY production system (right)(237). Production systems are colour coded in the key. Farms UK6, UK7 and UK 11were sampled on multiple occasions, permitting analysis of temporal variation, where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated. Multiple barns sampled for UK6 are indicated by letters (a) barn 1 (b) barn 2 and (c) barn 3. Median-joining tree analysis was completed on Network 10.0.0.0. A total of 145 positions were used in this analysis, encoding 29 nucleotide sequences.*

6.5 DISCUSSION

A total of 145 SNP markers were analysed via Mid-Plex genotyping for 75 separate pools of ~50-80 *D. gallinae* mites collected across the UK and Europe. A distinct population structure was observed including a high genetic diversity, defined by multiple unique haplotypes, together with notable linkage disequilibrium.

6.5.1 SNP annotation

The primary output of the GATK pipeline, as described in Chapter four, was a basic VCF file. The VCF lacked information relating to the context of the sequence that contained the variant; any potential effect on protein code (synonymous, missense or nonsense) or structure; association with phenotypes or diseases; tissue expression data or information relating to the likelihood that the variant is damaging. A number of bioinformatic tools for SNP annotation already exist (e.g. SNPit (596), SNPnexus (597), SNP Function portal (598), SNPper (599), Fans (600), FunctSNP (601), Annovar (602)). Presently, only a small number of tools are available for analysis of non-human SNP data (e.g. Fans, FunctSNP, Annovar, VEP). Many tools that are more general can only analyse species with SNP information in dbSNP and some require that SNPs being annotated already exist in dbSNP (603). Analysis of *D. gallinae* data with these tools proved unsuccessful due to the nature of the draft genome assembly available and lack of annotated SNPs. In order to achieve SNP annotation, after selection of the target SNPs, the .gff file associated with the draft *D. gallinae* was searched to provide information relating to location in an exonic or intronic region of the genome and identification of any known associated proteins. As SNPs were identified through comparison of transcriptomic *D. gallinae* data with the draft *D. gallinae* genome, SNPs were expected to be located in exons. Overall, 61% were identified in exons, 14% were mapped to candidate introns, and 25% were not found to have any annotation in the available .gff file (Figure 42). Mapping of markers to intronic regions could be the result of genomic DNA contaminants in the samples used to generate the original transcriptomics data, as well as inaccurate curation of current draft genome annotation on intron/exon boundaries. In total, 16% of SNPs were in genes with homologues annotated in other organisms, with the closest including related species such as *Ixodes scapularis* (Table 30). The SNPs selected from vaccine candidates were identified to be in exons, but no corresponding gene features were identified in the .gff file of draft genome annotation. Investigation into the 25% of SNPs where annotation could not be found demonstrated no clear difference in the read depth or quality score when plotted against each other (Figure 43). Despite the absence of current annotation, no difference in mapping quality was observed (Figure 43), indicating that absence of annotation would not impact on SNP genotyping success and highlighting that current genomic resources

are not fully curated at the time of writing. It is recognised that the assembly and annotation remain works in progress.

6.5.2 Biological validation of the GATK pipeline

There are a number of ways to biologically validate computationally derived SNPs, as discussed in chapter four (2.5.2), including Sanger sequencing (491), high-throughput SNP genotyping methods including SNP assays (487, 488), high-resolution melting (HRM) (489) and mass spectrometry (490). Sanger sequencing of ten SNP markers (i.e. 10% of the primary SNP marker panel) was conducted (using (UK6, UK11, UK9, ROM2, BEL1, POR2)) as preliminary biological validation before SNP genotyping was completed to ensure that the GATK pipeline had produced variants that were likely to be polymorphic across populations (Table 29). Two sites failed to amplify successfully, but due to time constraints it was not possible to allocate additional time to PCR optimisation. Seven markers showed allelic diversity across the samples and one demonstrated only the alternative sequence type (Mite_2307). SNP genotyping results for this marker revealed the alternative allele present in 50 samples, with the reference allele present in 25 samples, confirming that the lack of diversity observed through Sanger sequencing was due to the limited number of samples tested (i.e. six pooled populations). The combination of Sanger sequencing and SNP genotyping of 145 markers provides evidence that the self-validating GATK pipeline was computationally robust and the markers identified are valid for assessing genetic diversity across *D. gallinae* populations.

6.5.3 Genetic diversity

Average genetic diversity across all haplotypes was 0.3252 (Table 34). Differences in genetic diversity across geographic clusters was observed, with cluster three (France, Italy, Portugal and Spain) demonstrating the highest diversity. It should be noted that this could potentially be attributed to bias in sampling, with a higher number of farms included for Italy and Portugal compared to other countries sampled. Focusing on production systems, minimal difference was observed in genetic diversity between free-range and intensive systems, both across Europe and within the UK (0.0059 and 0.0121, respectively), with free-range being higher in both cases. The UK accounted for most free-range farms sampled in the European comparison, with six farms from Romania and one Albanian farm being the only other farms from Europe to utilise a free-range or rural backyard system.

6.5.3.1 Genetic diversity across production systems

Questionnaire data from the UK, and production system information collected during sample collection across the rest of Europe, indicated whether samples were collected from a free-range

and intensive production system origin. Genetic diversity across production systems was similar, both across the rest of Europe, with a difference of 0.0059, and across the UK, with a difference of 0.0121 (Table 34). Haplotype numbers were high in both production systems with no differentiation in phylogeny between system observed in the UK (Figure 45). This prompts rejection of hypothesis two, that there will be significant differences in genome-wide diversity across production systems. Free-range flocks have access to a larger environment for parasite transmission to occur (362) and higher *D. gallinae* populations have been reported on free-range farms compared to intensive systems (24, 26). Despite this, both production systems are subject to rapid mite population reduction during flock turn around, resulting from cleaning of poultry sheds, and suffer from unintentional mixing of new *D. gallinae* from infested equipment and/or hens.

6.5.3.2 Genetic diversity in the UK

Comparison between geographic clusters revealed that the UK samples presented higher diversity than cluster two (Belgium, Czech Republic, Denmark, Germany and the Netherlands), comparable diversity to cluster four (Albania, Greece and Romania), and lower diversity than cluster three (France, Italy, Portugal and Spain). In the UK, lower genetic diversity was observed in organic free-range samples when compared to non-organic free-range samples, with a difference of 0.0273. Organic farms are strictly limited in the use of chemical control measures and use of genetically modified crops (9), and as such control measures used against *D. gallinae* are predominantly desiccant dusts/silica powders. A greater selection pressure placed on *D. gallinae* treated by acaricidal or chemical control on non-organic farms could be one reason for the increased genetic diversity observed. Typically, selection pressure from acaricidal or chemical control would be expected to reduce genetic diversity, however, questionnaire data from the UK revealed a broad range of control measures have been adopted, including chemical control, desiccant dusts, natural control measures and hygiene related measures. Additionally, differing UK farmers used a range of combinations of control measures combined against *D. gallinae*. This could result in little or no directional selection on specific haplotypes across the UK as populations of *D. gallinae* as each farm are under specific selection pressures related to the chosen control measure in place.

6.5.4 Population structure

Results from phylogenetic and linkage analyses suggest a spatial structuring of genetic diversity in *D. gallinae* across Europe, with high haplotype numbers for populations and significant linkage disequilibrium. Spatial structuring of genetic diversity can occur when gene flow is not high enough for homogenisation of allele frequencies to occur throughout the studied area (604). As

well as gene flow and dispersal, genetic structure in populations can be influenced by genetic drift in small populations, which impacts genetic structure through increased differentiation (605). Due to *D. gallinae* residing off host and the sample population for this study predominantly focused on commercial laying farms (both intensive and free-range), populations of *D. gallinae* are likely to start from small founding populations bought onto farms via infested hens or equipment. Additionally, cleaning of poultry facilities between flocks of chickens can be expected to reduce *D. gallinae* population size, with questionnaire data from the UK showing 78% of farmers power wash poultry sheds as the preferred cleaning method. These founder events would further reduce effective population sizes and lead to higher genetic differentiation in *D. gallinae* via genetic drift.

6.5.4.1 Linkage disequilibrium

Linkage disequilibrium, denoted as *D*, is the non-random association of alleles at different loci (606). If *D* is equal to 0 there is linkage equilibrium (LE), sharing similarities to the Hardy-Weinberg equilibrium (HWE), implying statistical independence (606). LE and HWE both indicate that alleles at different loci are randomly associated with each other, however LE differs from HWE as it is not established in one generation of random mating (606). Results of linkage analysis revealed significant positive (P<0.01) linkage disequilibrium in all *D. gallinae* analysed, except for intensive UK farms (Table 34). It should be noted that UK intensive farms were only represented by three samples and this low number could be a reason for lack of significance. Overall, 82 markers demonstrated significant positive LD, 30 of which were confirmed after Bonferroni correction representing 20.7% of the 145 markers included in the panel (Table 35). Positive linkage disequilibrium shows that these 30 SNPs appear more frequently in combination with other markers than expected under neutrality theory. A subdivision in populations or changes in population size, as well as the exchange of individuals across populations, can all affect LD through a genome (606). Intentional or unintentional mixing of individuals from subpopulations with differing allele frequencies creates LD in the genome (607, 608). As previously stated, unintentional mixing of *D. gallinae* occurs when infested transport/equipment or hens are bought onto farms and this can occur within individual countries from suppliers or across countries through trade links. Under recombination alone, the decay of LD can be distorted but if selection maintains differences in allele frequencies across two or more loci among subpopulations then LD will persist in each subpopulation (609, 610). An increase in LD can also occur when changes in population size occur, especially when extreme reductions occur, such as a population bottleneck. Whilst a 'natural' population of *D. gallinae* (i.e. residing outside of commercial poultry systems) might be expected to adhere to HWE, populations of *D. gallinae* sampled from commercial poultry houses, as studied here, will

likely be subject to migration and selection that will cause the HW assumptions to be rejected. Bottleneck events result in changes in gene frequency and increases in LD (by inducing gene frequencies amongst loci), but also result in gene frequencies correlations within individual loci that result in a deviation of genotypic frequencies from HWE, that can influence genetic variance (611). Consistent bottlenecks in *D. gallinae* populations residing on farms occurs during flock rotation, due to rapid reductions in population size during cleaning of hen houses, followed by population expansion occurring when new hosts (i.e. a batch of hens) become available for feeding to occur. A combination of unintentional mixing and frequent extreme population reductions due to cleaning of poultry houses between flocks could explain the significant LD observed in *D. gallinae* populations studied. Technical factors, including the locations of markers selected relative to each other, could potentially play a role in the LD observed.

6.5.4.2 Phylogenetic relationships from Network analysis

6.5.4.2.1 Spatial analysis across Europe

Network analysis showed 75 distinct haplotypes for all *D. gallinae* populations sampled (Figure 44) with no shared haplotypes observed. Clustering of samples from Romania and Belgium suggests less differentiation between *D. gallinae* populations in these countries when compared to Portugal and the UK, where haplotypes are seen spread across the network map with close phylogeny to multiple countries. Farms sampled in Romania and Belgium were from a limited geographic range, especially when compared to the size of the country in the case of Romania (see Figure 7, Chapter three), which could explain the close phylogenetic relationship observed. As genetic drift can impact genetic structure in small populations through increased differentiation (605), it is possible that small founding populations in farms have differentiated through genetic drift during population expansion and colonisation of individual farms, shifting allelic frequencies to form individual haplotypes. Whilst, trade across Europe of poultry, and poultry related equipment could result in the unintentional movement of *D. gallinae* populations, and thus introduction and mixing of new alleles, it could be an insufficient level of gene flow to permit homogenisation of allele frequencies to occur resulting in spatial genetic diversity observed across *D. gallinae* populations sampled (604).

Samples from UK farms were shown predominantly clustered together (15/29 farms), in close relationship with farms from Greece and Albania on the network map (Figure 44). In contrast, individual haplotypes from UK were shown to be phylogenetically related to farms from Belgium, Portugal, Romania, the Netherlands, indicating historical admixture of *D. gallinae* populations between Europe and the UK has occurred.

6.5.4.2.2 Spatial analysis across the United Kingdom

Across the UK, two haplogroups were observed, with phylogeny linking farms from England, Wales and Northern Ireland, suggesting admixture of *D. gallinae* populations across the UK occurs. Samples taken from UK6 across three poultry barns revealed distinct haplotypes for each barn sampled, suggesting differentiation between the populations of *D. gallinae* residing in each barn. An organic free-range production system was utilised with desiccant dusts as the primary control measure in place and high hygiene standards (personally observed over four years of sampling). Two barns sampled from UK14 also demonstrated distinct haplotypes with clear phylogenetic differentiation between populations, with the haplotype from one barn in haplogroup A and the haplotype from the other barn found in haplogroup B (Figure 45). Distinct populations could possibly be the result of effective control measures and high hygiene measures preventing unintentional mixing of populations, producing three separate subpopulations residing on the same farm. In order to confirm this, SNP genotyping of multiple pooled *D. gallinae* samples from each barn would be required to provide further clarification and evidence to support this notion.

6.5.4.2.3 Temporal analysis across the United Kingdom

Distinct haplotypes were observed for all farms with temporal samples (UK6, UK7 and UK11), indicating changes in the genetic diversity of *D*. *gallinae* populations over time. Different patterns of phylogeny were observed for each farm, implying that changes in diversity are also related to individual farms and do not follow a universal pattern. Samples from UK11, spanning from 07.12.2018 to 29.05.2019 were shown to have a close phylogenetic relationship, clustered in haplogroup A. Samples taken from UK6 indicate temporal changes in the genetic diversity of *D. gallinae* populations over the years of 2017-2020 with distinct haplotypes observed for each time point sampled. A close phylogenetic relationship was observed between three samples collected from the same barn close together in time (six months apart; 14, 15 and 20 months after initial sampling, respectively), with greater phylogenetic separation seen from the initial time of sampling (Figure 45). Through questionnaire data collected at the first time point (+0, Figure 45), flock age was known to be 1.3 years old at sampling. This means that a change in flock occurred between the first time point (+0) and subsequent time points (+14, +15, +20), providing one plausible explanation for phylogenetic differentiation. Changes in population size during flock turnaround could result in changes in allele frequencies over time and thus the appearance of new haplotypes. Different haplotypes observed in the same barn from close time points could reflect different subpopulations following expansion, influence from cleaning between flocks or unintentional mixing of new *D. gallinae* into the farm.

Interestingly, two samples from UK7 were clustered in haplogroup A (from 30.07.2017 and 16.09.2018), whilst one sample (from 30.09.2017) was most closely related to UK6 in haplogroup B. Despite two months difference between two time points, clear phylogenetic separation was observed in the *D. gallinae*, which could represent two, or more, subpopulations of *D. gallinae* in the same flock Under optimal conditions, *D. gallinae* can complete its lifecycle within ~7-10 days (20, 32), so population expansion can occur very quickly, with rapid shifts in allelic frequencies theoretically possible.

Results from UK6, UK7 and UK 11 show clear indications of changes in *D. gallinae* over the course of time, meaning that hypothesis three (that genetic diversity of populations will remain constant over time) should be rejected. Despite this, a greater number of *D. gallinae* pools or individual *D. gallinae*, from each time point, a greater number of time points and increased number of farms sampled over time would be required to provide a deeper understanding in the changes in *D. gallinae* populations.

6.5.5 Considerations and complications faced in SNP genotyping *D. gallinae* samples

Due to limitations surrounding the availability of genomic data relating to *D. gallinae*, the status of the draft *D. gallinae* genome assembly and DNA extraction limitations, there were a number of complications and considerations faced during SNP genotyping.

6.5.5.1 Base calling quality

One characteristic of sequence reads which needs to be taken into consideration is base-calling scores. During sequence information generation, a probability value is assigned to a nucleotide for a certain position, typically called a “PHRED score”, named after the base-calling software tool Phred (18, 19). The phred score can be calculated by the equation –10 log10 P (the base calling is false), and typically a score of 20 is considered a reliable threshold, equivalent to a false-positive rate of 1% (456). This was factored in SNP selection for the SNPs identified in *D. gallinae* and SNPs included in the genotyping assay has a PHRED scale ranging from 28.07 to 2552.41.

6.5.5.2 DNA concentrations from individual mites

The required concentration per sample for Eurofins Mid-Plex genotyping is 25ng/µl, however this was unachievable when extracting DNA from an individual *D. gallinae*. Nanodrop readings from individual mites included in the first place sent for genotyping ranged from 1.961 to 55.208 ng/µl (average 11.698 ng/µl) for 20µl. Test PCR amplifications using these samples were successful (demonstrated by successful amplification of amplicons from the COI (Chapter 4) and

Vitellogenin (Chapter 7) loci), however upon quantification at Eurofins using a Qubit Analyser, accurate DNA concentrations were sub-optimal. After assay optimisation, amplification of samples with a concentration of 0.1 ng/µl or higher was achieved by Eurofins but overall, many samples failed to pass quality control and could not be called successfully. This required a revised approach where DNA was extracted from pooled *D. gallinae* per each farm, or existing single preparations were pooled, in order to achieve higher DNA concentrations. Subsequently, DNA concentrations were vastly improved with the analysed samples ranging from 1.000-141.592 ng/µl (average 22.261 ng/µl) when quantified on a Qubit analyser.

6.5.5.3 Success rates for multiplex PCR

Eurofins Mid-Plex genotyping method involves multiplex PCR requiring multiple primer pairs. Typically, PCR amplification of a single SNP locus will result in over 85% successful amplification with SNP scoring identifying a genotype for 95% of samples (612). Amplification of multiple SNPs together commonly reduces the success rate to 50-70% of SNPs and the amount of product generated can vary greatly, between 10 and 1000 folds (612). Calling rate for samples can also decrease, resulting for example in SNPs that are scored for sample A failing to be scored for sample B or C, meaning that samples can have incomplete genotype data (612). Calling rate can be influenced by the quantity of DNA provided, with lower concentrations reducing call rate. In combination with the variable DNA concentrations provided for SNP genotyping, this can help explain why some SNP markers failed to amplify at all and samples failed to successfully call sufficient genotypes to be included in analysis. After consultation with Eurofins regarding DNA normalisation, it was agreed that no normalisation would be undertaken at the RVC due to the previous complications. If any normalisation would be required, it would be completed by Eurofins.

6.5.5.4 Allele frequency estimation

Data generated from pooled samples sequenced at high depth can be used for direction estimation of population allele frequencies based on the relative abundance of reads with an alternative allele (613, 614). However, obtaining equimolar amounts of DNA from every individual included in the pool can be challenging due to stochastic variation in the efficiency of amplification of individual's DNA. This can result in a bias of the occurrence of different alleles causing low confidence in allele frequency estimates (548). In an ideal situation, heterozygous individuals would have a 50/50 distribution of reference and alternative alleles but in reality the data is skewed one way or another, due to a number of reasons (486). In biological terms, allele-specific expression patterns, where one allele (e.g. reference) is more highly expressed compared to the other (e.g. alternative), can potentially throw off genotype estimates (486). At

an individual level, this issue is unlikely to cause serious effects as the expression bias would need to be several orders of magnitude for inaccurate calling of homozygote or heterozygote. However, when pooled samples are sequenced, small differences in expression can result in throwing off allele frequency estimates (486). For linkage and phylogenetic analysis, all heterozygote samples were converted to the dominant allele present in the population, reducing the impact of allele frequency estimation errors in the results.

6.6 CONCLUSION

Analysis of 145 SNP markers from 75 pooled *D. gallinae* samples collected from across the UK and Europe showed high spatial genetic diversity, with no conserved haplotypes detected at more than one location or on more than one occasion. Significant linkage disequilibrium was detected across all populations with the exception of intensive layer farms, where sample size may have been limiting, indicating historical and on-going admixture between *D. gallinae* populations.

7 PYRETHROID RESISTANCE IN *DERMANYSSUS GALLINAE* AND FREQUENCY IN UK AND EUROPEAN POPULATIONS

7.1 GENERAL INTRODUCTION

In agriculture, the control of arthropod pests remains heavily dependent on acaricidal or insecticidal application (615). The habitual and intensive use of acaricides and insecticides has resulted in the widespread occurrence of resistance development in over 500 arthropod species (616, 617). For a number of important livestock parasites, crop pests, disease vectors and urban pests, resistance has developed to such an extent that their control has become challenging (617).

7.1.1 Resistance to acaricides

Resistance of an arthropod population to a pesticide is typically defined as a heritable decrease in susceptibility to that pesticide, leading to inadequate field control (www.irac-online.org/about/resistance/). Resistance is agreed to be an evolutionary phenomenon, with the same factors driving the dynamics of resistance to classical chemical pesticides and to biopesticides (618). Insecticide and acaricide resistance is a major threat for the chemical control of insects and mites in public health and agriculture (619). At present, the Insecticide Resistance Action Committee (IRAC) distinguishes between at least fifty-five different chemical classes and more than twenty-five distinct mode of action (MoA) groups (620). MoA diversity is of key importance for effective Insecticide Resistance Management (IRM). However, the costs involved in the discovery, development and marketing of chemicals with new properties have increased immensely in recent years and slow down the development of compounds with new MoA, in conjunction with increased regulations surrounding chemical use and licensing. In addition, concerns about environment and human health, integrated in new regulations, demand molecules with better selectivity (621). To preserve the utility and diversity of available and newly developed insecticides/acaricides, it is critical to understand the resistance mechanisms that mediate against these compounds (620) and develop diagnostic tools that support monitoring activities and resistance management.

Currently, one emphasis in resistance research involves unravelling underlying molecular mechanisms in order to control the development and spread of resistant populations (617). Understanding the molecular mechanisms can help manage resistance through identification of specific changes occurring at the genomic level, and subsequently in development of robust diagnostics. Meanwhile, the characterisation of detoxification enzymes (enzymes involved in detoxification metabolism) aids development of add-ons as well as 'resistance-breaking' compounds to be used in insecticide formulations (617). Compared to insect pests, knowledge on the resistance mechanisms of Acari of major economic important has not kept pace, although progress has been made (617).

7.1.2 The development of resistance and mechanisms of resistance

Resistance to acaricides/insecticides is influenced by a multitude of factors, including ecology, biology, control operations and genetics (117), with a range of adaptations that allow insects or mites to overcome lethal doses (617). These adaptations can be classified, based on biochemical or physiological properties, as either mechanisms of decreased exposure (excretion, metabolism, distribution or penetration) or decreased response to acaricides (interaction with its target site) (118, 119). Most cases involve changes in target site sensitivity resulting from point mutations (pharmacodynamic mechanisms), or as a result of metabolism and/or sequestration of the acaricide before it can reach the target site due to qualitative or quantitative changes in critical detoxification enzymes (pharmacokinetic mechanisms) (123, 617, 618, 622). Detoxification enzymes including P450 monooxygenases, glutathione-S-transferases and esterases, have been reviewed elsewhere (118, 120, 122, 123).

When resistance is polygenic, that is resistance controlled by a combination of multiple genes and their products, the overall level of resistance can be attributed to the sum of individual gene contributions (623, 624), although antagonistic or synergistic interactions may occur (625-627). The relative contribution that each individual resistance locus provides to complex acaricide/insecticide resistance phenotypes has only been sporadically researched (628). Resistance mechanisms in mites and insects can be complicated, and the relative contribution of target-site mutations in relation to resistance phenotypes is not always known (619). Specifically, the relative contribution and importance of target site mutations can be difficult to assess by association of phenotype with mutation occurrence in field populations as prolonged selection may lead to an accumulation of multiple resistance mutations (619). In addition, the majority of research focusing on resistance levels and/or epistatic interactions confirmed by a single genetic factor can be challenging to interpret if the resistance alleles are not investigated with a common genetic background (627, 629-632). As such, where possible, analysis of resistance traits requires that the strains being studied should be identical except for the causal gene (633, 634).

7.1.3 Use of acaricides to control *D. gallinae* and resistance mechanisms

Control of *D. gallinae* chiefly depends upon acaricide applications and relatively few are approved for the purpose worldwide. Organophosphates, carbamate, amidine and pyrethroid-based acaricides are the most widely used. Furthermore, many acaricides are not specifically labelled for use against *D. gallinae* and if not properly applied, the development of acaricide resistance can be accelerated (19, 108, 114-116).

7.1.4 Pyrethrins/Pyrethroids for the control against *D. gallinae* and resistance mechanisms

Pyrethrins are naturally occurring compounds that are derived from members of the Chrysanthemum family (635). Against arachnids they have a fast knock-down effect, but they are unstable in the environment and may not always remain active long enough to effectively kill mite populations (635). Pyrethroids are a synthetic adaptation of pyrethrins that are designed specifically for longer stability and therefore have a longer lasting effect in the environment (636). Pyrethrins and pyrethroids work by acting as neurotoxins on motor and sensory nerves of the central nervous and neuroendocrine systems of insects. Pyrethroids are lipophilic, that is a tendency to combine or dissolve in lipids or fats, which aids in their efficacy as contact insecticides (635). Their mode of action involves blocking sodium ion movement along the axon of the nerve fibre (637) causing repetitive nerve discharges that result in paralysis and death (638).

The principle genetic difference contributing to pyrethroid resistance is considered to be conserved point mutations in the voltage gated sodium channel gene (639), but metabolic resistance mediated by P450s or carboxylesterases are also well documented (640, 641). In cattle ticks, the involvement of P450s, glutathione S-transferases and esterases have been linked to pyrethroid resistance (642).

7.1.4.1.1 Voltage-gated sodium channel (VGSC)

For pyrethroids, the target site is the voltage-gated sodium channel (VGSC) (639). It has been demonstrated for some insect pests that the VGSC is encoded by two genes (known as *VGSC1* and *VGSC2*), both have a similar function (643), however this duplication is not commonly witnessed in arthropods (635). In mite and insect species, enhanced enzymatic detoxification and target-site insensitivity are the predominant mechanisms of resistance associated with pyrethroids (617, 622), typically referred to as knockdown resistance (*kdr*) (644). Multiple studies have demonstrated that *kdr* and *kdr*-type resistance is due to point mutations in the *para* family of voltage-gated sodium channel genes (644). In *T. urticae* the major enzymes involved in pyrethroid resistance have been demonstrated to be carboxylesterases (CarE) and cytochrome P450 monooxygenases (645-647). Overall, more than 50 sodium channel polymorphisms or combinations have been linked to resistance to pyrethroids in mites, insects and ticks (648), with a number of them functionally validated through expression in *Xenopus* oocytes (618, 649). The majority of mutations in the VGSC have been located in transmembrane segments 4 to 6 in domain II (referred to as IIS4-IIS6), including changes at residues M918, in the IIS4-IIS5 linker, as well as L925, L923 and T929 in the IIS5 segment, and L1014 in segment IIS6 (639, 650). In the S6 transmembrane segment of domain III a resistance mutation (F1528I) has

also been demonstrated (651) and through functional assays has been shown to confer strong sensitivity to multiple pyrethroids (652, 653).

7.1.5 Implications of acaricide resistance

Due to difficulties in controlling *D. gallinae* and differing levels of resistance demonstrated to multiple acaricides (measured by the lethal concentration required), a greater understanding of resistance mechanisms and frequency of resistant genotypes present in *D. gallinae* populations is essential (264). Genome plasticity permits for rapid selection and dissemination of polymorphisms with a selective advantage to a given parasite and for *D. gallinae* this could mean developing further resistance mechanisms to new or existing drugs (264).

7.2 BACKGROUND INFORMATION

Work by Katsavou et al., (53) focusing on identification of pyrethroid resistance in *D. gallinae* underpins the work reported in this chapter and informed sequencing and SNP genotyping for mutations.

7.2.1 Identification of target site mutations associated with pyrethroid resistance (completed by collaborators)

VGSC domain II (S4-S5-S6) and III (S6) sequences were obtained from the recently available draft *D. gallinae* genome sequence (232). None of the previously characterised pyrethroid resistance mutations were present in domain IIS4-S5 and IIIS6 of the VGSC sequenced from a susceptible strain SUSC (Sample GER1 in this PhD thesis, relating to a population of *D. gallinae* kept on hens at the University of Hannover that have not been subjected to acaricide selection). Katsavou et al., performed comparison of two Greek populations (GRC1 and GRC2) collected from the field, previously shown to confer resistance through bioassays, with the SUSC strain, detecting non-synonymous mutations codons 918, 925 (Domain II) and 1534 (domain III) in both Greek populations (*M. domestica* numbering; deduced amino acids shown in Table 36, with functional validation in *Xenopus* labelled). Mutations at these positions have been associated with pyrethroid resistance in other species (Figure 47). Mutations at M918L and F1534L appeared fixed in the GRC1 population, while M918L and L925V appeared fixed in the GRC2 population, which demonstrated the strongest resistance phenotype. They defined the term 'fixed' as manual curation of sequencing chromatographs showing no evidence of an alternative nucleotide at the position of interest, based on sequencing PCR fragments amplified from pooled material. They noted that this measure does not necessarily mean complete fixation at

the population level, as a high number of individual mites would need to be assayed to provide proof.

Amino acid position*	Nucleotide triplets and amino acids identified in the PRM VGSC			
M918	ATG (M)	TTG/CTG (L)	GTG (V)	ACG (T)
L925	CTG (L)	GTG (V)	ATG (M)	
T929	ACC (T)	ATC (I)	TTC (F)	
I936	ATC (I)	TTC (F)		
F1534	TTC/TTA/TTG (F)	CTG/CTC (L)		
F1538	TTC (F)	CTC (L)		

Table 36: *Nucleotide triplets present in the sequenced VGSC gene segments of D. gallinae mites. The nucleotide triplets and amino acids identified in the GRC1 and GRC2 strains that were tested with bioassays are underlined. Nucleotide triplets/amino acids found in the SUSC strain are indicated in normal font, while those found in the screened populations are indicated in bold font. * indicating* Musca domestica *numbering. Table reproduced from (592).*

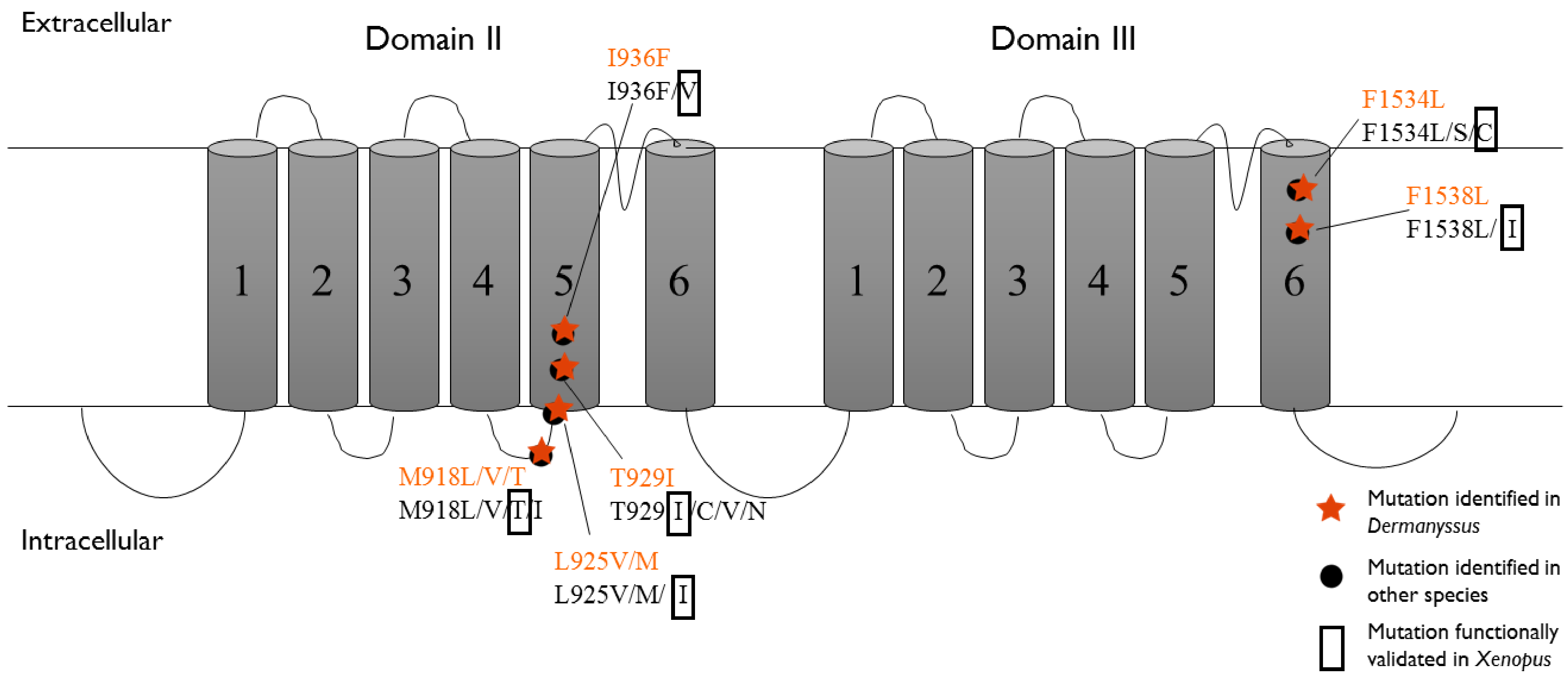


Figure 47*: Schematic diagram of domain II and III of the VGSC. Mutations found in D. gallinae are indicated with orange circles, while those found in other species are indicated with black circles. Those mutations that were functionally characterized in* Xenopus *are framed in a box (649). Figure reproduced from (592).*

7.2.2 Aims and hypotheses

Aim: The main aim of this chapter was to assess the occurrence of VGSC genotypes that have been previously related to resistance to pyrethroid acaricides

Hypothesis 9: Genetic markers that have been functionally linked to acaricide resistance will be detected in European populations of *D. gallinae.*

7.3 METHODOLOGY

7.3.1 Sample selection

A total of 53 *D. gallinae* populations were sampled in the first-round study (Figure 48). Samples from Germany, Belgium and Greece were provided by collaborators, samples from all other locations were as described in the General Methodology (Sections 2.1.1-2.1.3).

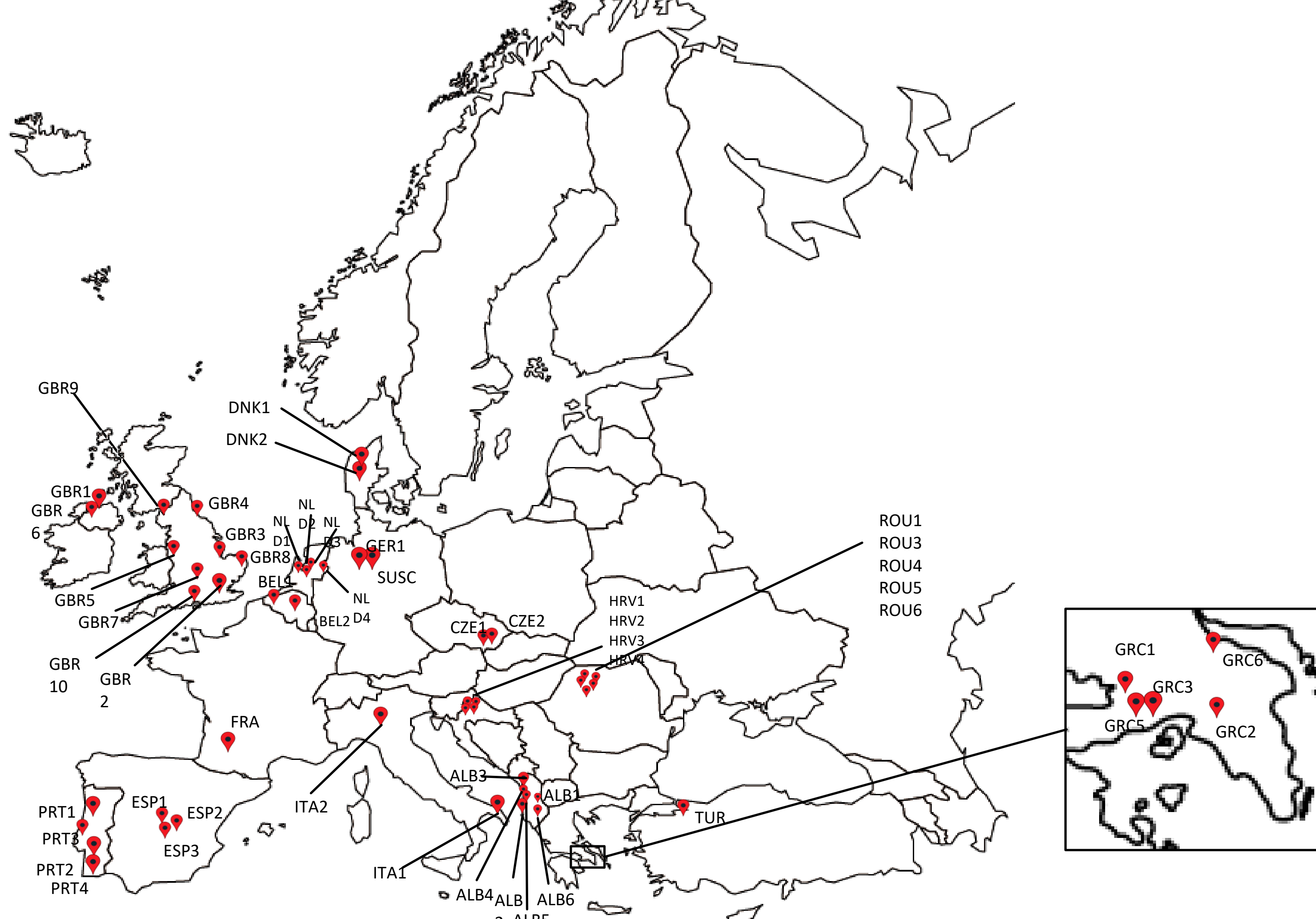


Figure 48: *Map showing the origin of 53 D. gallinae populations analysed in this study, spread across 15 European countries, where GBR: United Kingdom, DNK: Denmark, FRA: France, PRT: Portugal, ITA: Italy, ALB: Albania, TUR:*

Turley, ROU: Romania, HRV: Croatia, CZE: Czech Republic, GER: Germany, BEL: Belgium, NL: the Netherlands and GRC: Greece.

7.3.1.1 United Kingdom

From the UK a total of ten farms were included, eight from England and two from Northern Ireland (Table 37).

Country	Paper allocated no.	Farm no.	County	Production type
Northern Ireland	GBR1	UK22	Tyrone	Free-range
England	GBR2	UK6	East Sussex	Free-range
England	GBR3	UK14	Lincolnshire	Free-range
England	GBR4	UK1	Cheshire	Free-range
England	GBR5	UK8	Shropshire	Intensive
Northern Ireland	GBR6	UK23	Tyrone	Free-range
England	GBR7	UK5	Kent	Free-range
England	GBR8	UK11	Suffolk	Free-range
England	GBR9	UK4	Cheshire	Intensive
England	GBR10	UK12	West Sussex	Free-range

Table 37: *UK farms sampled for the analysis including county, farm no. allocated in PhD thesis, farm name allocated for publication and the production system utilised by the farm.*

7.3.1.2 DNA extraction

Genomic DNA was extracted from pools of approximately 50 mites per sample according to General Methodology 2.3 (Routine DNA extraction from whole mites) at the RVC. For the collaborator's samples, DNA was extracted using DNAzol reagent (Molecular Research Center, Inc., Ohio, USA) according to the manufacturer's instructions.

7.3.1.3 Primer design (completed by collaborators)

Two fragments of the VGSC locus were amplified: domain IIS4-S5 (145bp) and domain IIIS6 (237bp). For the amplification of IIS4-IIS5 the following primers were designed: KF1 5'-CAAGTCATGGCCGACGTTGA-3' and KR1 5'-GTCGGTGTAGTTCTTGCCGAA-3', and for the amplification of IIIS6 the following primers were designed: KF3 5'-AAAAGACGACCAGCCCGACT-3', KR3 5 -GGTATGGCTTTGGCGGGTTT-3'. Primers were used with a final concentration of 2mM. Primers were designed based on the VGSC gene sequence from the recently published PRM genome (232) and used to PCR amplify partial sequences of the *D. gallinae* VGSC gene.

7.3.1.4 PCR and Sequencing

PCR, PCR purification and Sanger sequencing was completed as per General Methodology sections 2.7 Polymerase chain reaction (PCR), 2.10 PCR Purification and 2.11 Sanger Sequencing for samples completed at the RVC. Samples provided by the collaborator were processed and tested elsewhere.

7.3.2 SNP genotyping of M918, L925 and F1534 mutations

Sequences relating to previously identified knockdown resistance mutations (kdr), *kdr* 1 and *kdr* 3 (i.e. mutation at M918, L925, and F1534), were provided by collaborators after initial identification in Greek populations (Section 7.2.1). Information regarding positions 918, 925 and 1534 was annotated (Figure 49) and used for inclusion here in the SNP genotyping panel for application in a wider, second round study. SNP markers were designed as detailed in Chapter 6 (6.3.3.1.2), based on the sequences in Figure 49.

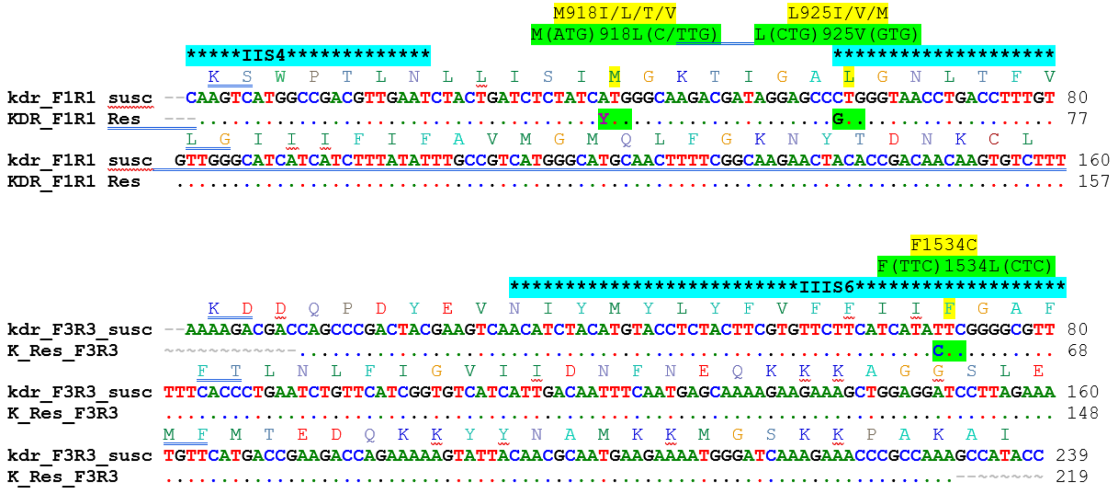


Figure 49: *Nucleotide and amino acid sequence of KDR1 and KDR3 from D. gallinae with pyrethroid mutations annotated M918, L925 and F1354 codon positions.*

7.4 RESULTS

7.4.1 Sanger sequencing of pyrethroid mutations of *D. gallinae* populations from UK and Europe

Fifty-three samples from 15 European countries were examined for the presence of M918L, F1534L and L925V, as well as possible additional pyrethroid resistance mutations (Figure 47), by direct sequencing of PCR products representing two fragments of the VGSC gene, covering domains II (IIS4-IIS5) and IIS6. Mutations previously associated with resistance were detected as double peaks (defined here as not fixed in the population) or single peaks (defined here as fixed in the population) in sequence traces, indicating the presence of a mixture of resistant and susceptible alleles or fixed resistant/susceptible alleles in each sample, respectively, subject to the caveats outlined above (Section 7.2.1).

With the exception of the susceptible strain (SUSC, from Germany), and the samples GER1 (same population as SUSC, but different DNA extraction) and UK12, all samples analysed (50/53) had at least one mutation at an amino acid position previously associated with pyrethroid resistance (Table 38). Overall, eight putative pyrethroid resistance mutations were identified across all *D. gallinae* populations at positions M918 (918L, 918V, 918T), L925 (925M, 925V), L929 (925V and 925M) and I936 (936F) in domain II, with two additional mutations, F1534L and F1538L in domain III (Figure 47).

7.4.1.1 M918 substitutions

The M918L substitution was the most commonly occurring mutation, identified in 37 out of 53 samples (69.8%) with the amino acid leucine found to be the predominant allele. In contrast, the amino acid valine was identified at codon 918 (M918V) in only one out of 53 samples (NLD2, Netherlands). The amino acid threonine (M918T) was also rare, found in four out of 53 samples (7.5%), with three from UK farms (UK4, UK8, UK14).

7.4.1.2 Other substitutions

L925V was also frequently detected in 19 out of 53 samples (35.8%), in contrast to methionine at the same position (three out of 53 samples). In general, the rarest substitutions were T929I and I936F, each found in only four (7.5%) out of 53 samples. Two additional mutations were found in domain III, F153L and F1538L (Table 38). F1534L was the most frequent mutation, as it was present in 31 out of 53 samples (58.5%). F1538L was found only in three samples (5.7%): two derived from UK (UK4, UK14) and one from France (FRA3).

7.4.1.3 Country level

Focusing on the country level, *D. gallinae* populations from three countries – Greece, Croatia and Albania – appeared to have the same VGSC gene mutation profile, as in all of these populations M918L, L925B, and F153L were present (Figure 50). Samples originating from Portugal (PRT1, PRT2, PRT3, PRT4) and Spain (ESP1, ESP2, ESP3) also had a similar profile, as they all showed a combination of M918L with F1534L, with the only difference being fixation at F1534 in Portuguese samples. The population collected from France (FRA) had the M918T and F1538L substitution. Additionally, in populations from Italy (ITA1, ITA2) we identified two mutations, M918L and L925V. The sample from Turkey (TUR) was the only population bearing the T929I mutation alone. T929I was also present in Romania (ROU1, ROU3, ROU4, ROU5, ROU6), along with M918L and F1534L. In samples from Czech Republic (CZE1, CZE2) M918L, L925M/V and I936F were present. In the Netherlands, all three most commonly found mutations were identified – M918L, L92V and F1534L – while M918V was found in only one sample. Belgian samples appeared to have M918L, L925M, I936F and F1534L. The samples obtained from Denmark (DNK1, DNK2) harboured both M918L and L925V.

7.4.1.3.1 United Kingdom

Compared to the other European countries sampled, the UK had the highest rates of fixed mutations present, with four sites demonstrating fixation in at least one population (Figure 50), suggesting a high level of pyrethroid resistance across the UK. The highest diversity of mutations was found in UK samples: M918L/T, L925V/M, I936F, F1534L and F1538L. In addition, in two UK samples (UK5, UK11) a point mutation was identified (TTT to CTT, not fixed in the population) at the 1537 position, corresponding to a phenylalanine to leucine substitution.

Focusing on the four positions in domain II, no UK farm had a mutation present at site T929 (Table 38, Figure 50). Position M918 was more polymorphic compared to position L925, with 80% of UK farms having a mutation at position M918 compared to 30% at position L925 (Table 38). UK14 showed the greatest variation in amino acid substitutions with two alleles at both M918 and L925, 30% of UK farms showed two alleles at one position, 50% showed one non-wild type allele and 20% had the wild-type allele (Table 38).

Code	Country	Region	Domain II				Domain III	
			M918	L925	T929	I936	F1534	F1538
SUSC	Germany		M	L	T	I	F	F
GER1		Hannover	M	L	T	I	F	F
GRC1	Greece	Megara	L	L	T	I	L	F
GRC2		Ilion	L	V	T	I	F	F
GRC3		Megara	L	V	T	I	F	F
GRC5		Megara	L	V	T	I	F	F
GRC6		Avlona	L	V	T	I	F	F
BEL1	Belgium		M/L	L/M	T	I/F	F/L	F
BEL2			L	L	T	I	L	F
PRT1	Portugal	Zezerovo	L	L	T	I	L	F
PRT2		C	L	L	T	I	L	F
PRT3		Riveria	M/L	L	T	I	F/L	F
PRT4		Casa de Hounance	M/L	L	T	I	F/L	F
GBR1	UK	Dungannon-Ireland	L	V	T	I	F	F
GBR2		Eastsussex	L	L	T	I	L	F
GBR3		Lincolnshire	T/L	L	T	I	F/L	F/L
GBR4		West(Durham)	L	L/V	T	I	F/L	F
GBR5		Shropshire	M/T	L/M	T	I	F	F
GBR6		Dungannon-Ireland	M/L	L	T	I/F	F	F
GBR7		Woodlands	M	L	T	I/F	F	F
GBR8		Suffolk	L	L	T	I	F/L	F
GBR9		Peter house farm	T	L	T	I	F	L
GBR10		WestSussex	M	L	T	I	F	F
ALB1	Albania	Peshkopi	L	L/V	T	I	F/L	F
ALB2		Lushnye	L	L/V	T	I	F/L	F
ALB3		Shkoder	L	L/V	T	I	F/L	F
ALB4		Durres	L	L/V	T	I	F/L	F
ALB5		Berat	L	L/V	T	I	F/L	F
ALB6		Korca	L	L/V	T	I	F/L	F
CZE1	Czech Republic	South Moravian region	M/L	L/V	T	I	F	F
CZE2		Bohemia	M	L/M	T	F	F	F
TUR	Turkey	Karacaali	M	L	I	I	F	F
ESP1	Spain	Castile-LaMancha	L	L	T	I	F/L	F
ESP2		Castile-LaMancha	L	L	T	I	F/L	F
ESP3		Castile-LaMancha	L	L	T	I	F/L	F
HRV1	Croatia	Zagreb	M/L	L	T	I	F/L	F
HRV2		Zagreb	M	L	T	I	F/L	F
HRV3		Zagreb	M/L	L	T	I	F/L	F
HRV4		Zagreb	M/L	L/V	T	I	F/L	F
NLD1	Netherlands	ODNK	M/L	L/V	T	I	F/L	F
NLD2		BONW-BROCH	M/V	L	T	I	F/L	F
NLD3		A. Van de Braak	M/L	L	T	I	F/L	F
NLD4		Aalten	M	L	T	I	F/L	F
FRA	France	Grenade	M/T	L	T	I	F	F/L
DNK1	Denmark	Vejle	M/L	L/V	T	I	F	F
DNK2		Jylland	M/L	L/V	T	I	F	F
ROU1	Romania	Tatarlaua	L	L	T	I	L	F
ROU3		Cuzdrioara	M	L	T/I	I	F	F
ROU4		Cuzdrioara	M	L	T/I	I	F/L	F
ROU5		Cuzdrioara	M	L	T/I	I	F/L	F
ROU6		Floresti	M/L	L	T	I	L	F
ITA1	Italy	Lecce	M/L	L/V	T	I	F	F
ITA2		Cremona	L	V	T	I	F	F

Table 38: *Amino acid substitutions in two VGSC domains (II and III) of European D. gallinae populations. The susceptible alleles are indicated with bold font. A forward slash, separating amino acids, indicates that the allele is not fixed in a certain* D. gallinae *populations. Table reproduced from (592)with names corresponding to paper publication.*

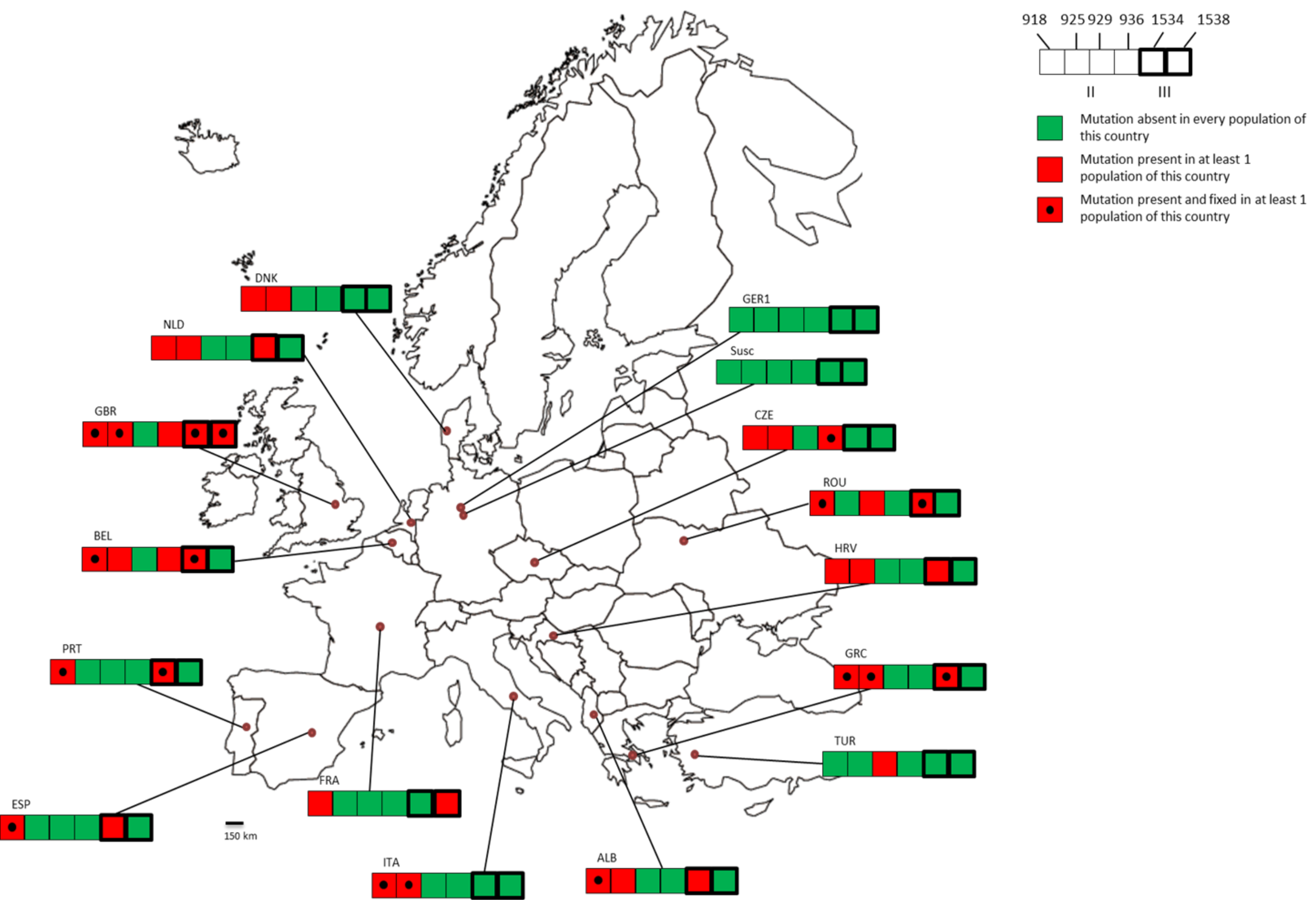


Figure 50*: Schematic illustration of the distribution of pyrethroid resistance mutations in D. gallinae populations across Europe. The classification of mutations was based on visual inspection of sequencing chromatographs and comprised three categories: 'absent', 'present' and 'fixed' (when no background signal was detected at the investigated position). Figure reproduced from (592) with names corresponding to paper publication*

7.4.2 Mid-plex genotyping of pyrethroid mutations of *D. gallinae* populations from UK and Europe

Ninety-two pooled *D. gallinae* samples from 70 farms, 52 from Europe and 18 from the UK, were genotyped for the M918, L925, T929, I936, F1534 and F1536 mutations previously identified (Table 39-40). Eighty-eight samples (96%) had at least one mutation present relating to pyrethroid resistance, with four samples (4%) demonstrating genotypes associated with susceptibility at all sites (ALB3, ALB4, GER1 and ITA17) (Table 40).

In total, nine putative pyrethroid resistance mutations were identified across all *D. gallinae* populations at positions M918 (918L, 918V, 918T), L925 (925M, 925V), T929 (929V and 929M) and I936 (936F) in domain II, and three additional mutations, F1534L, F1537L and F1538L in domain III (Figure 47).

7.4.2.1 M918 substitutions

The M918 substitution was the most frequent mutation, occurring in 86 out of 92 samples (93.5%) with the amino acid leucine (M918L) found to be the predominant allele. In contrast, the amino acid valine (M918V) was identified in seven samples (7.6%) with three of these found in populations from UK6 on different occasions over time. The amino acid threonine (M918T) was demonstrated in eight out of 92 samples (8.7%), with four from UK farms (UK7.1, UK8.1, UK9, UK10 and UK11).

7.4.2.2 Other substitutions

In domain II, L925 substitutions were detected in 27 out of 92 samples (29.3%), where valine substitutions were detected in 25% of populations in contrast to methionine at the same position (9.8%; Tables 39-40). The rarest substitution was T929I, found in only seven *D. gallinae* populations (7.6%) from four countries (France, the Netherlands, Romania and the UK). I936F substitutions were found in 28 (29.3%) out of 92 samples.

Mutations at three additional sites were found in domain III, F1534L, F1537L/S and F1538L (Table 39-40). F1534L and F1537L/S were both present in 35 out of 92 samples (38.5%). At F1537L/S, the predominant change was to leucine (34 samples) and to serine in just one sample (UK14.1). In contrast, F1538L was found in 14 samples (15.2%).

7.4.2.3 Country level

Across Europe, individual VGSC genotype profiles (II and III domains) were observed for all countries sampled, with the UK demonstrating the only profile to have a mutation at all sites (Figure 51) and the highest rate of fixation. Germany was the only country represented by a

sample defined by no mutations at any site, due to the population sampled originating from a susceptible strain of *D. gallinae* housed in a closed population at the University of Hannover. Fixation of at least one mutation was observed in nine of the countries sampled, with no fixation seen in the Czech Republic, Germany, Greece and the Netherlands (Figure 51). A phenylalanine to leucine substitution was the most common substitution seen at F1354, but in one population from France (FRA3) and two populations from Italy (ITA14, ITA16) a phenylalanine to cysteine substitution was revealed (Table 40).

		918	925	929	936	1534	1537	1538
Whole dataset	Homozygous							
	Susceptible	6.52%	70.65%	92.39%	69.57%	57.78%	57.78%	81.11%
	Resistant	30.43%	4.35%	0.00%	1.09%	3.33%	3.33%	0.00%
	Heterozygous							
	Susceptible + Resistant	59.78%	23.91%	7.61%	29.35%	35.56%	34.44%	15.56%
	Resistant	3.26%	1.09%	0.00%	0.00%	0.00%	0.00%	0.00%
UK	Homozygous							
	Susceptible	0.0%	69.4%	94.4%	61.1%	80.00%	34.29%	82.86%
	Resistant	25.0%	0.0%	0.0%	0.0%	0.00%	8.57%	0.00%
	Heterozygous							
	Susceptible + Resistant	72.2%	27.8%	5.6%	38.9%	20.00%	57.14%	20.00%
	Resistant	2.8%	2.8%	0.0%	0.0%	0.00%	0.00%	0.00%
Europe	Homozygous							
	Susceptible	11.5%	76.9%	98.1%	80.8%	47.1%	78.4%	86.3%
	Resistant	36.5%	7.7%	0.0%	1.9%	5.9%	0.0%	0.0%
	Heterozygous							
	Susceptible + Resistant	55.8%	23.1%	9.6%	25.0%	49.0%	21.6%	13.7%
	Resistant	3.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 39: *Summary of resistant and susceptible VGCS (domain II and III) allele occurrence in D. gallinae sampled from the UK and the rest of Europe.*

Sample	Domain II				Domain III		
	918	925	929	936	1354	1537	1538
ALB1	**L**	L	T	I	F	F	F
ALB3	M	L	T	I	F	F	F
ALB4	M	L	T	I	F	F	F
ALB5	M	L	T	I	F	F/**L**	F
ALB6	M/**L**/**V**	L/**V**	T	I	F/**L**	F	F
BEL1	M/**L**	L/**M**	T	I	F/**L**	F	F
BEL2	M/**L**	L/**M**	T	I/**F**	F/**L**	F	F
BEL3	**L**	L	T	I	**L**	F	F

BEL4	**L**	L	T	I	**L**	F	F
BEL5	M/**L**	L	T	I/**F**	F/**L**	F/**L**	F
BEL6	M/**L**	L	T	I/**F**	F/**L**	F	F
CZH2	M/**L**	L/**V/M**	T	I/**F**	F/**L**	F	F
DEN1	M/**L**	L/**V/M**	T	I	F	F	F
DEN2	M/**L/T**	**V**	T	I	F	F	F
FRA1	M/**L**	L/**V**	T/**I**	I/F	F	F	F
FRA2	M	L	T	F	F	F	F
FRA3	**L**	L/**V**	T	I/**F**	F/**C**	F	F/**L**
FRA4	M/**L**	L/**V**	T	I/**F**	F	F	F/**L**
GER1	M	L	T	I	F	F	F
GRC1	M/**L**	L	T	I	F	F/**L**	F/**L**
GRC2	M/**L/T**	L	T	I	F	F/**L**	F/**L**
GRC3	**M/L**	L	T	I	F	F/**L**	F/**L**
GRC4	**M/L**	L/V	T	I	F/**L**	F/**L**	F
ITA10	M/**L**	L/V	T	I	F	F	F
ITA13	**L**	L	T	I	F	F	F
ITA14	M/**L**	V	T	I	F/**C**	F	F
ITA15	**L**	V	T	I	F	F	F
ITA16	M/**L**	V	T	I	F/**C**	F	F
ITA17	M	L	T	I	F	F	F
NET7	M/**L**	L	T	I	F	F/**L**	F
NET8	M/**L**	L	T/**I**	I/**F**	-	-	-
NET9	M/**T**	L/**V/M**	T	I/**F**	F	F/**L**	F/**L**
POR1	**L**	L	T	I	F/**L**	F/**L**	F
POR2	M/**L**	L	T	I/**F**	F/**L**	F	F
POR3	**L**	L	T	I	F/**I**	F	F
POR4	L/**V**	L	T	I	F/**L**	F	F
POR5	M/**L**	L	T	I	F/**L**	F	F
POR6	M/**L**	L	T	I/**F**	F/**L**	F	F
POR7	**L**	L	T	I	F/**L**	F/**L**	F
POR8	**L/V**	L	T	I	F/**L**	F/**L**	F/**L**
POR9	**L**	L	T	I	F/**L**	F/**L**	F
POR10	**L**	L	T	I	F/**L**	F	F
POR11	**L**	L	T	I	F/**L**	F	F
ROM1	**L**	L	T	I/**F**	F	F	F
ROM2	**L**	L	T	I/**F**	F	F	F
ROM3	M/**L**	L	T/**I**	I	F/**L**	F	F
ROM4	M/**L**	L	T/**I**	I	F/**L**	F	F
ROM5	M/**L**	L	T/**I**	I	F/**L**	F	F
ROM6	**L**	L	T	I	F/**L**	F	F
SPA1	**L**	L	T	I	**L**	F	F
SPA3	**L**	L	T	I	F	F	F
SPA4	**L**	L	T	I	F/**L**	F	F

Table 40: *Amino acid substitutions in two VGSC domains (II and III) of European D. gallinae populations. The susceptible alleles are indicated with bold font. A forward slash, separating amino acids, indicates that the allele is not fixed in a certain D. gallinae population*. An – indicating samples where data was unavailable to call.

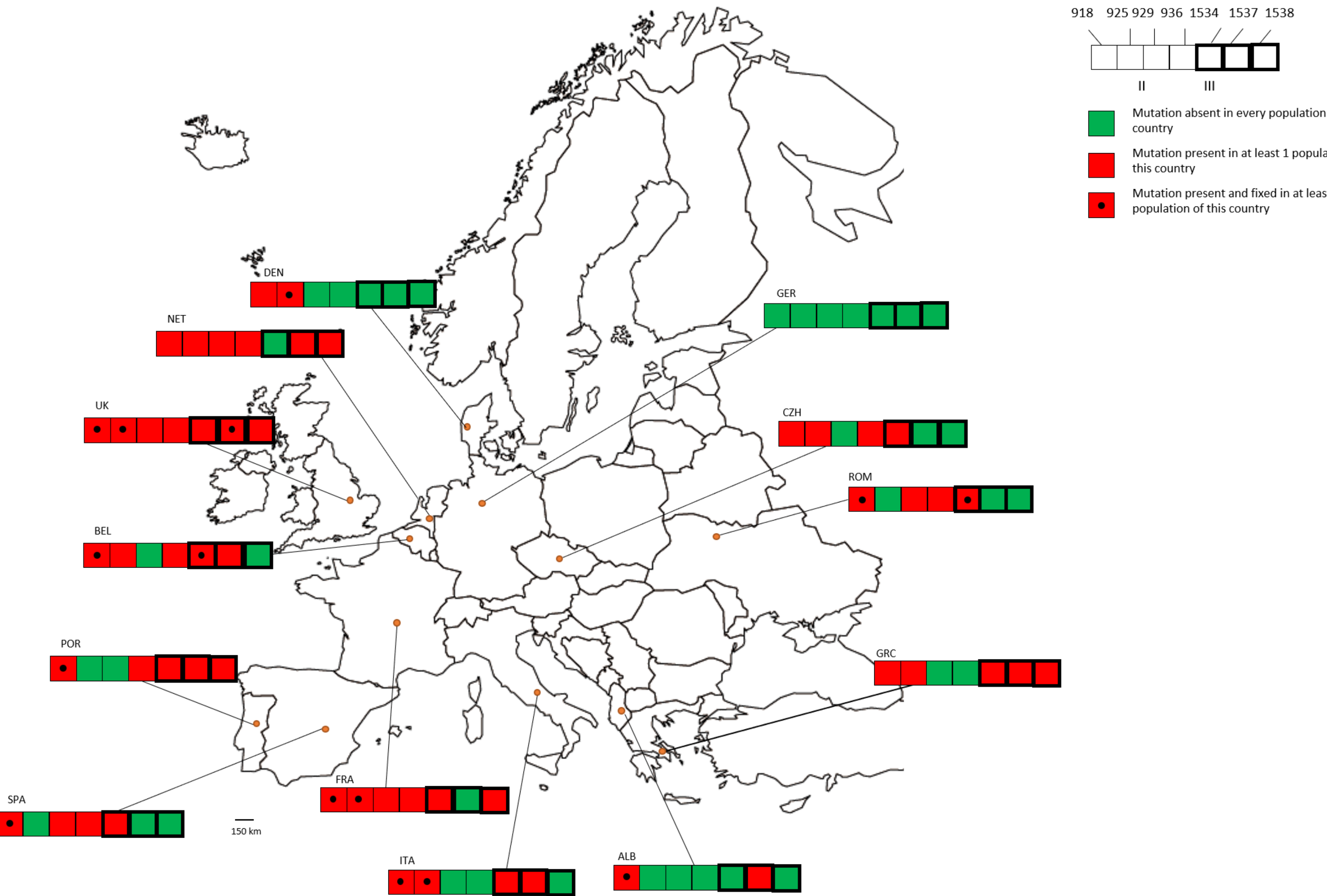


Figure 51: *Schematic illustration of the distribution of pyrethroid resistance mutations in D. gallinae populations across Europe from SNP genotyping. The classification of mutations was based on presence or absence of multiple alleles comprised to make three categories: 'absent (reference only allele), 'present in 1 population' (heterozygous reference and alternative alleles or heterozygous alternative alleles) and 'fixed in 1 population' (alternative only alleles present).*

7.4.2.4 United Kingdom

A total of 36 pooled *D. gallinae* samples were genotyped from 18 farms across the UK, with all 36 farms showing at least one mutation related to pyrethroid resistance (Table 41, Figure 52). Resistance related mutations were observed in all four countries: England, Wales, Scotland and Northern Ireland (Figure 52). The most polymorphic site was M918, with all 36 samples showing a mutation present and 25% of populations samples showing fixation.

Mutations in domain II at L925 were observed in 11 samples (30.5%) and at I936 for 14 samples (38.9%), with substitutions at T929 the rarest with just two samples (5.6%), demonstrating presence of alternative alleles (UK3 and UK6.12) (Table 41). In domain III, the F1537 mutation was the more frequently discovered, with 65.7% (8.6% homozygous) of samples demonstrating a substitution at this position compared to 21.6% of European populations. One sample, UK14.1 showed a phenylalanine to serine substitution, whilst all other samples were phenylalanine to leucine (Table 41).

7.4.2.4.1 Samples collected over time

Temporal changes in mutations relating to acaricide resistance were observed at three UK farms with samples collected at different time points (Table 41). Changes in both occurrence and number of substitutions were seen following multiple sampling events from UK6, UK7 and UK 11.

For UK6, variation between barns was also noted, including polymorphism at M918 (barns a and c), L925 (barn c), T929 (barn c), and I936 (barns a,b and c) for domain II, and at F1534 (barns a and b), F1537 (barn a) and F1538 (barn a). For UK7, changes were observed at M918, L925, I936 and F1537, whilst for UK11 changes were only seen at M918 and F1357 (Table 41).

Sample	Domain II				Domain III		
	918	925	929	936	1534	1537	1538
UK1	**L**	L	T	I	F	F/**L**	F
UK2	**L**	L	T	I	F	F	F
UK3	M/**L**	L	T/**I**	I	F/**L**	F/**L**	F
UK5	M/**L**	L	T	I/**F**	F	F/**L**	F/**L**
UK6.1 (a+0)	**L**	L	T	I	F	F/**L**	F
UK6.2 (a+1)	M/**L**	L	T	I/**F**	F	F/**L**	F
UK6.3 (a+0)	**L**	L	T	I	F/**L**	F/**L**	F
UK6.4 (a+4)	**L**	L	T	I	F	**L**	F
UK6.5 (a+6)	L/**V**	L	T	I	F	F/**L**	F
UK6.6 (b+6)	M/**L**	L	T	I	-	-	-
UK6.7 (c+6)	**L**	L	T	I	F	F	F
UK6.8 (a+14)	**L**	L	T	I	F	**L**	F
UK6.9(a+14)	M/**L**	L	T	I/**F**	F	F/**L**	F
UK6.10(a+20)	M/**L/V**	L	T	I	F	F/**L**	F/**L**
UK6.11(b+20)	M/**L**	L	T	I/**F**	**F/L**	F/**L**	F
UK6.12(c+20)	M/**L/V**	L	T/**I**	I/**F**	F	F	F
UK7.1 (+0)	M/**L/T**	L/**V**	T	I	**F/L**	F/**L**	F
UK7.2 (+2)	M/**L**	**M/V**	T	I	**F/L**	F	F
UK7.3 (+14)	M/**L**	L/**V**	T	I/**F**	**F/L**	F/**L**	F
UK8.1	**T**	L/**M**	T	I	F	F	F
UK8.2	M/**L**	L	T	I/**F**	F	F/**L**	F/**L**
UK9	M/**L/T**	L/**M/V**	T	I	F	F/**L**	F/**L**
UK10	M/**L/T**	L/**V**	T	I/**F**	F	F	F
UK11.1 (+0)	M/**L**	L	T	I	F	F/**L**	F/**L**
UK11.2 (+10)	M/**L**	L	T	I/**F**	F	F/**L**	F/**L**
UK11.3 (+15)	M/**L/T**	L	T	I	F	F/**L**	F/**L**
UK12	M/**L**	L	T	I	F	F/**L**	F
UK13	**L**	L	T	I	**F/L**	F	F
UK14.1 (A)	M/**L**	L/**V**	T	I	F	**S**	F
UK14.2 (B)	M/**L**	L/**V**	T	I/**F**	F	F/**L**	F
UK15	M/**L**	L	T	I	F	F/**L**	F
UK18	M/**L**	L	T	I/**F**	F	F	F
UK20	M/**L/V**	**L/M/V**	T	I/**F**	F	F	F
UK22.1	M/**L**	L/**V**	T	I	F	F	F
UK22.2	M/**L**	L/**V**	T	I/**F**	F	F	F
UK23	M/**L**	L	T	I/**F**	F	F	F

Table 41: *Amino acid in two VGSC domains (II and III) of UK D. gallinae populations. The susceptible alleles are indicated with bold font. A forward slash, separating amino acids, indicates that the allele is not fixed in a certain D. gallinae. An – indicating samples where data was unavailable to call. Letters a-c indicating barns relating to the same farm. Farms UK6, UK7 and UK 11were sampled on multiple occasions, permitting analysis of temporal variation, where the date of visit is indicated by +0 for the first visit and the subsequent gap in months indicated.*

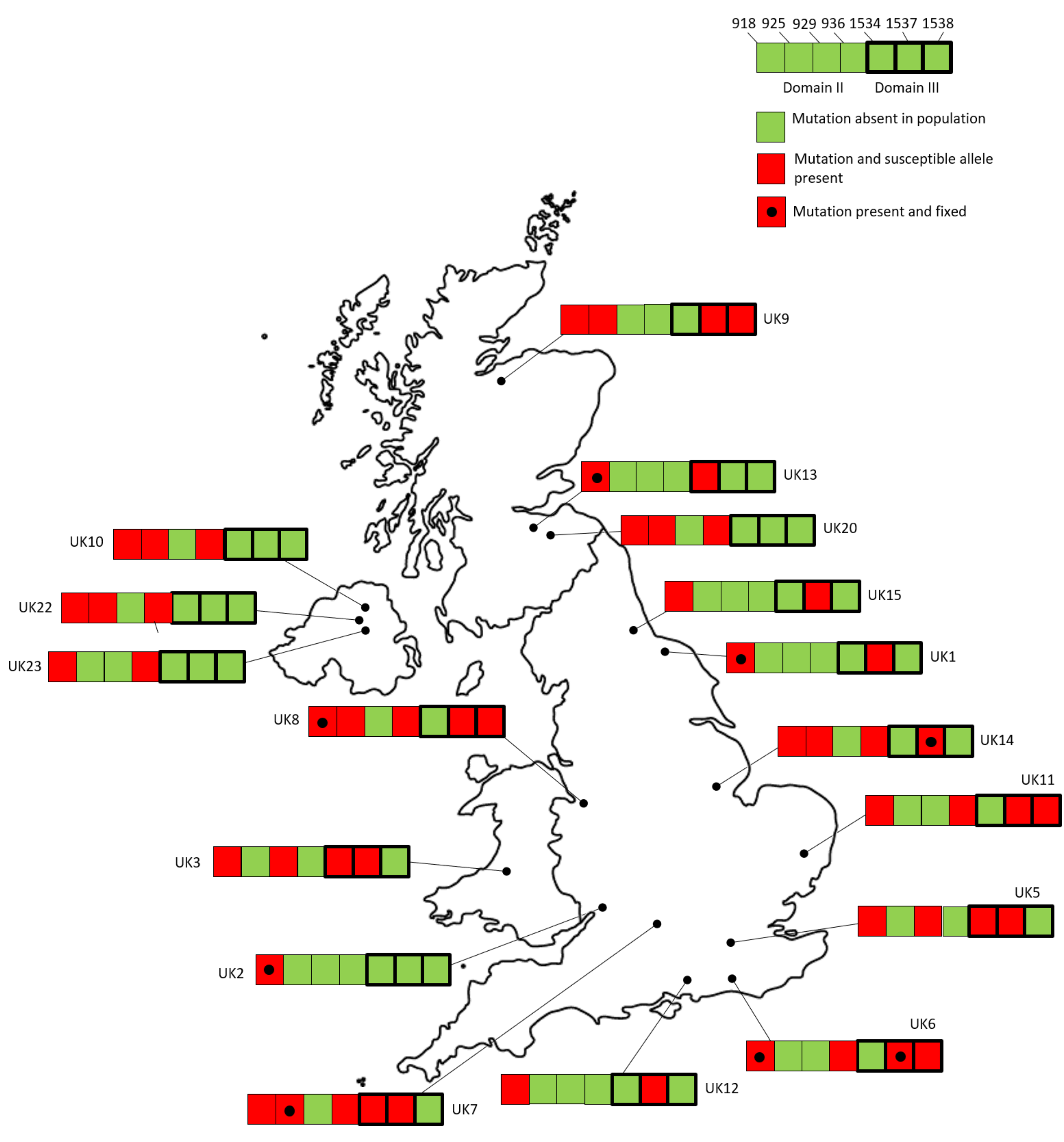


***Figure 52**: Schematic illustration of the distribution of pyrethroid resistance mutations in D. gallinae populations across the UK from SNP genotyping. The classification of mutations was based on presence or absence of multiple alleles comprised to make three categories: 'absent (reference only allele), 'present in 1 population' (heterozygous reference and alternative alleles or heterozygous alternative alleles) and 'fixed in 1 population' (alternative only alleles present).*

7.5 Discussion

A combination of Sanger sequencing and Mid-plex SNP genotyping in two overlapping studies revealed the occurrence of nine mutations putatively associated with pyrethroid resistance across multiple *D. gallinae* populations at positions M918, L925, L929, I936 in domain II of the VGSC gene, and three additional mutations, F1534L, F1537L and F1538L in domain III.

7.5.1 Resistance in UK populations

Results demonstrated the UK had the highest occurrence of sequence types associated with pyrethroid resistance of all countries analysed, with four polymorphic sites demonstrating fixation in at least one population (i.e. homozygous resistant allele only observed) (Figure 50). Two of the UK farms sampled utilised intensive production systems (UK4, UK8), whilst the remaining eight all used free-range systems, but there appeared to be no obvious difference in acaricide resistance between production systems. SNP genotyping revealed a higher occurrence of resistant genotypes in mite populations from UK farms compared to European farms for all polymorphic sites, except for F1534 (Table 39) with substitutions found at all nine polymorphic sites across the UK. Fixation was observed at three sites (M918, L925 and F1537) but was not demonstrated at F1354, as seen by Sanger sequencing (Figures 50-51). Fixation (i.e. homozygous allele only) was seen at one site for five UK farms, and two sites for one farm (UK6). Despite the high occurrence of markers associated with resistance seen in the UK, one UK farm had no mutations present in any of the examined domains from Sanger sequencing, UK12 (GBR10), implying no resistance to acaricides in the *D. gallinae* population on this farm. Questionnaire data showed that UK12 is a free-range, organic farm which utilises no acaricides in the control of *D. gallinae*. From personal experience the farm has a very small number of chickens (~100 per flock), housed in trailers converted to hen houses. At the time of sampling no live *D. gallinae* were located, only a smaller number of dead *D. gallinae* mites found on metallic support beams on the barn. However, subsequent SNP genotyping revealed substitutions present at M918 and F1537 for UK12, although substitutions relating to pyrethroid mutations were present at very low levels, 5% and 10% of read coverage, respectively.

In most examples VGSC profiles were individual to farms sampled across the UK (Figure 52). Exceptions were UK1 and UK15, where similar VGSC profiles (II and III) were observed, and UK10 and UK22 from Northern Ireland, where identical VGSC profiles were observed (Figure 52). This could be due to proximity of these farms geographically, when compared to other UK farms, or shared suppliers for UK10 and UK22. Shared suppliers might include areas such as chicks, food or bedding supplies, farm equipment etc. UK23 shared the same supplier as UK10 and UK22, but the polymorphism at L925 was not detected (Table 6). Farms located nearby in Scotland (UK13,

UK20) presented different marker profiles. Variation in VGSC profiles could result from multiple independent selection events relating to resistance development or stemming from high levels of pre-existing diversity in dynamic and complex populations, such as *D. gallinae*. This is especially true in cases where no single marker was subjected to preferential selection, for example when all individual markers confer compatible levels of resistance to pyrethroids. Questionnaire data collected from the UK revealed a range of combinations of control measures in place at the time of sampling (Chapter 3). Three farms genotyped used an intensive production system (UK8, UK13, UK15) and 15 were free-range, with five of these organic (UK3, UK6, UK9 UK20, UK23). Organic farms are not permitted to use chemical control measures in the control of *D. gallinae* populations, so polymorphisms present in populations sampled from these farms is likely due to VSGC profile of the founding *D. gallinae* population or unintentional admixture with resistant populations.

7.5.1.1 The M918T mutation

Sanger sequencing from UK4 revealed that the M918T and F1538L mutations were fixed in the *D. gallinae* population. UK4 reported no use of chemical treatment for the control of *D. gallinae* on their farm in response to the questionnaire, relying on the use of desiccant dusts and detergents, but noted that they had previously used organophosphates (although there is no link to M918 at this time). The M918T mutation is also known as the *super-kdr* mutation (knock-down resistance) and has been reported in *Tetranychus evansi* and a multitude of other pyrethroid-resistant arthropod populations (649, 654). The presence of this super-kdr mutation could be indicative of a more intense and broader resistance phenotype in the UK, as it occurs more frequently than other countries sampled. SNP genotyping revealed all populations (100%) sampled from the UK had M918 mutations, with 25% of populations exhibiting a fixed mutation (Table 38), indicating higher prevalence than European farms were 11.5% of populations had a susceptible phenotype.

7.5.1.2 The pyrethroid sensing residue: F1357

Sanger sequencing revealed that two farms (UK4 and UK14)) possessed a point mutation at position 1357 that resulted in a phenylalanine to leucine substitution. Of these farms, UK14 utilised a free-range system whilst UK4 operated an intensive system. Subsequently, SNP genotyping revealed a broader presence of mutations at F1537 (Tables 38, 40). Across the whole dataset, fixation of this mutation was observed in 3.3% of samples, but heterozygosity was seen in 34.4% of samples, however occurrence was higher in the UK compared to the rest of Europe. In the UK 65.7% (8.6% homozygous) of samples demonstrated a substitution at this position compared to 21.6% of European populations. Two distinct barns were SNP genotyped from

UK14, confirming the point mutation observed by Sanger sequencing in barn B (phenylalanine to leucine), however in barn A a point mutation causing a phenylalanine to serine substitution was observed (Table 40). This was the only phenylalanine to serine substitution observed across all 92 UK and European populations genotyped at F1357. Research has previously characterised this mutation as the 'pyrethroid sensing residue', also known as F3i16 (655), however it has not yet been validated as a functional mutation associated with resistance. However, given its position directly next to the major pyrethroid resistance mutation F1538I in both tick and mite VGSC genes it is hypothesised that it could play a role in resistance to pyrethroids.

7.5.1.3 F1538 mutations

In domain III, a F1538L mutation that has previously been demonstrated to have an association to pyrethroid resistance in the pollen beetle *Meligethes aeneus* (656) was detected in two UK farms (UK4 and UK14) and one French farm. UK4 and UK14 are located in two different counties across the UK, Lincolnshire and Cheshire (~150 miles apart), utilising two different production systems (one free-range and one intensive) indicating no clear connection between the two *D. gallinae* populations. Questionnaire data showed that UK4 reported no current use of acaricides at the time of sampling, UK14 responded to the questionnaire identifying chemical control as their only control measure against *D. gallinae*. SNP genotyping showed the F1358 mutation to be present in seven of the 36 UK samples analysed, with all samples showing a heterozygous mixture of resistant and susceptible alleles.

A similar mutation, F1538I, has been confirmed to play a role in pyrethroid resistance through electrophysiological studies (653), and demonstrated to confer a very high level of resistance to multiple pyrethroids in both tick and mite species (652, 657, 658). Introgression of F1538I into a susceptible *T. urticae* population revealed a strong pyrethroid resistance phenotype (619). As leucine and isoleucine amino acids are both branched-chain amino acids with similar structures and physiochemical properties, the resulting impact of the alternative mutation identified in *D. gallinae* is likely to be equal to the effect of F1538I. However, the role of this mutation remains to be functionally validated either alone or in combination with other mutations that have been identified in *D. gallinae*.

7.5.2 Presence and frequency of pyrethroid resistance mutations across Europe

Analysis of the presence and frequency of pyrethroid resistance mutations in a large number of *D. gallinae* populations across Europe through Sanger sequencing (53 samples in total from 15 European countries) was undertaken, aiming to investigate the presence and geographical distribution of M918L, F1534L and L925V, as well as possible additional mutations, in several

countries where pyrethroid resistance has already been reported (25, 108, 115, 635). Mutations associated with pyrethroid resistance were common across Europe, as only three out of 53 examined samples did not have any mutations in the VGSC domains (IIS4-S5 and IIIS6) that were investigated in this study. Subsequent SNP genotyping from a larger number of samples revealed 96% had at least one mutation present relating to pyrethroid resistance, with four samples (4%) examined showing no mutations in VGSC domains (ALB3, ALB4, GER1 and ITA17). Sanger sequencing revealed mutations at M918, L925 and F1354 for populations from ALB3 and ALB4, indicating that different subpopulations of *D. gallinae* were sampled for each study. M918L and F1534L, mutations that were associated with the striking pyrethroid resistance phenotype in the GRC1 and GRC2 populations from Greece (See Katsavou et al., (592) and Section 7.2.1), were the most common mutations across Europe (37 out of 53 and 31 out of 53, respectively). SNP genotyping also demonstrated M918 as the most common mutation in all samples analysed, with just 6.52% of samples showing no mutation present Table 39). Sanger sequencing revealed the M918L and F1534L mutations were detected in combination and both fixed in field populations from Portugal (PRT1, PRT2), UK (UK6) and Romania (ROU21). In other cases, the M918L and L925B mutations were, like, in the GRC2 population, both present and fixed in populations from Greece (GRC3, GRC5, GRC6), UK (UK22) and Italy (ITA2).

SNP genotyping of nine mutations putatively associated with pyrethroid resistance revealed a difference in prevalence of mutations across some countries (Figures 50-51). VGSC profiles based on eight sites (excluding F1537 as it was not included in the original data) were the same for Denmark, Belgium, Germany and Greece. In comparison, VGSC profiles for the Czech Republic, France, Italy, Spain, the Portugal, Netherlands and the UK all showed an increase in at least one population having a mutation present (Figure 51). Albania was the only country to have less mutations present in SNP genotyping data compared to Sanger sequencing, with no substitutions observed at L925 or F1534. Differences in VGSC profiles could result from different subpopulations captured in each *D. gallinae* pool from which DNA was extracted and analysed, due to higher sensitivity in the SNP assay.

7.5.3 Future investigations into acaricide resistance in *D. gallinae*

Further research is required to understand the exact role that the mutations targeted here play in relation to pyrethroid resistance. Research in other arthropod and acari species has shown that the effects of specific mutations on pyrethroid resistance phenotypes, singularly or in combination, can vary largely in relation to fitness, intensity and specificity (649, 659). Additionally, the possibility of non-target site mechanisms influencing the resistant phenotypes observed in *D. gallinae* cannot be ruled out. In a number of species, combination of cytochrome

P450 monooxygenase mediated metabolic resistance with target site resistance has been discovered (659). Use of the *D. gallinae* draft genome assembly (232) could help to facilitate understanding of metabolic resistance. Whilst a full understanding of detoxification pathways might not be achievable, identification of key candidate genes could be useful.

7.6 CONCLUSION

Overall, the results obtained demonstrate that genetic markers associated with pyrethroid resistance are widespread in European *D. gallinae* populations with the patterns of mutations and genotypes shared across several countries. The results draw attention to regulations surrounding the use of pyrethroids, and other acaricides, against the use of *D. gallinae*. Variation in the VGSC profiles may be indicative of high levels of pre-existing genetic diversity in *D. gallinae* populations and/or multiple independent mutation events contributing to resistance.

8 GENE-SPECIFIC GENETIC ANALYSIS: ASSESSMENT OF EXISTING GENETIC DIVERSITY AND SIGNATURES OF SELECTION AT LOCI ENCODING ANTI-PRM VACCINE CANDIDATES.

8.1 INTRODUCTION

Despite investment of millions of dollars and decades of research, development of vaccines against parasitic infections remains relatively unsuccessful, especially for ectoparasites (175, 660). Many factors have contributed to this, including the fact that parasitic infections are often chronic in nature. This chronic nature can be attributed to parasites frequently eliciting ineffective and/or inappropriate immune responses in their host or dampening the host immune system, resulting in prevention of an effective and/or robust immune response (176). Many parasites follow complex lifecycles that can complicate the process of developing efficacious vaccines. Additionally, many parasites employ immune evasion strategies such as molecular mimicry, antigenic variation and/or sequestration at both the individual and population level (177). For ectoparasites, the development of vaccines is even more difficult where direct interaction with the host may be limited to invasive feeding.

8.1.1 Vaccination against arthropods

For ectoparasites, vaccine research has commonly focused on peptidases, proteinases and their inhibitors (208). This is due to the fact ectoparasites typically require some degree of tissue penetration and/or destruction as part of their lifestyle, with haematophagous parasites (such as *D. gallinae*) typically controlling haemostasis. From a research perspective the biochemical study of peptidases, proteinases and their inhibitors can be relatively easy and the proteins are typically ubiquitous (208). Thus, peptidases, proteinases and their inhibitors are a practical choice for candidate anti-arthropod vaccine antigens. In arthropod species, vaccine targets have commonly been located in salvia or the midgut, including molecules that have key roles in arthropod feeding, vector capacity and/or reproductive fitness (661, 662). Additionally, antigens of arthropods have been related to pathogens of which they play a vectoral role in an attempt to reduce infection burden and occurrence of the vectored disease (180, 663). For example, Subolesin in ticks and Akirin (he Subolesin ortholog in insects) are both transcriptional regulatory factors known to impact expression of genes involved in arthropod innate immune responses and a number of cellular pathways in response to pathogens (180, 663). These antigens have shown potential for use in vaccines targeting several arthropod vectors, including mosquitos, soft and hard ticks, sand flies, sea lice and *D. gallinae* (as discussed in the general introduction) and for use against transmission and/or infection of tick-borne pathogens including *Babesia bigemina*, *Borellia burgdorferi, Anaplasma marginale* and *Anaplasma phagocytophilum* (reviewed in (180, 663)).

8.1.1 Antigenic diversity

One characteristic of parasitic organisms is their capacity for adaption to changes in their environment (387, 664). During an infection, a parasitic organism's survival is not only dependant on its aptitude for host colonisation but also its ability to successfully counteract the host's defence mechanism(s) (665). One of the major processes allowing evasion of the immune system by parasites is antigenic variation, permitting the persistence of a chronic infection despite an ongoing immune pressure (666).

In terms of vaccine design, antigenic variation by many pathogenic species is problematic, representing a major reason for the lack of control in some infectious diseases (667). In arthropods, tests of protein components of saliva as vaccine candidates revealed that antigenic polymorphism could be problematic in development of anti-salivary vaccines (668, 669). The majority of salivary proteins are subjected to selective pressure by their host's immune system, thus it can be hypothesised they will be highly polymorphic and encoded by rapidly evolving genes (670). High frequencies of gene duplications observed in the salivary gland transcriptome of haematophagous arthropods provide further support for this idea (671, 672). In one of the few arthropod molecules in which sequence polymorphism has been studied in detail – namely, the pituitary adenylate-cyclase-activating polypeptide (PACAP) receptor agonist (maxadilan) of *Lutzomyia longipalpis* – the extent of polymorphism between individual flies collected from different geographical locations was significant (673). This polymorphism represents a mechanism that improves the fitness of the vector under the selective pressure of an anti-maxadilan immune response(674).

8.1.1.1 Selection pressure on antigenic diversity

Balancing selection resulting from immune related selection pressure favours maintenance of genetic diversity, with low to medium allele frequencies within a population and balanced allele frequencies between populations (675). Genomic regions that are polymorphic or polymorphic sites exhibiting such patterns are theorised to be under balancing selection, such as incurred by immunity, and as such directly contribute to antigenic diversity. In some cases, where antigens are experiencing strong immune selection, clusters of alleles or similar alleles have been demonstrated across broad geographic ranges (676-681). In contrast, SNP and polymorphic sites with low minor allele frequencies can be representative of deleterious mutations experiencing purifying selection, or recent polymorphisms that could increase in frequency (682). In a parasite population, these polymorphisms will typically only be found in a very small proportion of individuals. As the major goal of diversity-covering vaccines is to encompass the maximum

number of haplotypes found, several groups have removed rare polymorphic sites from their population genetic analyses focusing on vaccine antigens (677, 679, 683).

8.1.1.2 Evaluation of haplotypes

Through utilisation of network and clustering analysis the relationships between haplotypes from various populations, as well as the distribution and extent of clusters of closely related haplotypes, can be investigated (682). As a result, the most distantly related alleles can then be chosen for more comprehensive analysis or for incorporation into a vaccine in an attempt to cover diversity (682). Research has shown that this approach can be successfully used to identify distinct clusters of alleles as the basis for serotype prediction (681, 684). These analyses permit identification of the most common and distinct haplotypes, providing the basis for selection of haplotypes representing a large proportion of the population-wide diversity for vaccine candidate antigens (682). The outcome of this type of analysis can aid in determining the feasibility of covering all known diversity of the target antigen, the number of required haplotypes to cover diversity in a vaccine, and the predicted efficacy of vaccine candidates (682). Through inclusion of vaccine analysis it is possible to provide a reference point for estimation of vaccine allele or serotype frequencies and the results from this can enable a diversity framework to be constructed to estimate strain-specific efficacy during vaccine clinical trials (685), and for determination of parameters of allele specific and cross-reactive responses (681, 686, 687).

8.1.1.3 Studying antigenic diversity in vaccine candidates to inform vaccine design and development

Work to identify vaccine candidates for *D. gallinae* has focused on measures of efficacy, including mite mortality and/or fecundity. Consideration of pre-existing antigenic diversity can be an important additional screening tool, offering opportunities to prioritise antigens likely to have the broadest efficacy. During vaccine design and development, population genetics studies are helpful to define the diversity of candidate antigens, detecting polymorphisms that could contribute to future immune escape due to antigenic diversity (678, 688, 689) and investigating the geospatial distribution of predicted genotypes (675, 680). To aid in the understanding of antigenic diversity that affects vaccine efficacy and identification of potential serotypes, the target gene (or gene region) that encodes the candidate antigen should be amplified, sequenced and population genetic analyses undertaken including determination of regions under balancing or other selection (682). To enable accurate estimation of natural allele frequencies, collection of samples that represent the natural parasite population in a defined geographic area is useful (682). Samples that have been collected in the same geographic area can be used to accurately estimate sequence diversity at the target for that region (682). The basis of the majority of

informative analyses of balancing selection is allele frequency, so one important consideration is obtaining a sufficient dataset of typically to capture allelic diversity (690, 691). Analysing a smaller number of sequences can cause incorrect diversity estimates, as alleles can be under-represented in small population samples, which skews allele frequencies and diversity estimates (682). Whilst repetitive regions can be included in population genetic analysis, impact on antigenic diversity is not as severe from expansion and contraction of repeat arrays when compared to amino acid changes (685, 692). Despite this, defining alleles on the basis of repeat number or when considering indels, presence of absence of a particular nucleotide sequence, can be used for prediction of whether these polymorphisms are modulated by immune selection (682). Research into the antigenic variability of current vaccine candidates that induce an increase in mite mortality in *D. gallinae* would help to provide further clarification on suitability for potential vaccine development.

8.1.2 Chosen vaccine candidates

Four known vaccine candidates against *D. gallinae* were selected for investigation of antigenic diversity: Tropomyosin, paramyosin, Cathepsin-D and Vitellogenin. At the time of selection, they were considered to be among the leading candidates for use as in vaccines against *D. gallinae*.

8.1.2.1 Vitellogenin

Vitellogenins are proteins serving as precursors of vitellins, the major egg yolk protein in many oviparous vertebrates and invertebrates (693, 694). Vitellogenin synthesis occurs primarily in fat cells in tissue-, sex- and stage-specific manners (694). Once released into the haemolymph, incorporation into developing oocysts occurs by receptor-mediated endocytosis (694-696). In higher oviparous vertebrates, multiple forms of vitellogenin have been discovered (697), with three types of vitellogenin protein demonstrated in the chicken (698). Evidence of multiple vitellogenin's has also been reported in Crustacea (699-701).

Host serum immunoglobulins, after traversing the gut epithelium, can enter the haemolymph of haematophagous arthropods, where they can interact with haemolymph components (such as vitellogenin) and membrane receptors involved in processes including the incorporation of vitellogenin by receptor-mediated endocytosis by developing oocytes (Sauer et al, 1994). As a result, vitellogenin and similar molecules (e.g. hemelipoglycoprotein (702)) have been investigated as potential vaccine candidates for use against haematophagous arthropods (186).

8.1.2.1.1 Vitellogenin in Arthropod species

From Arthropoda, a number of complete vitellogenin coding DNA sequences have been generated. In Crustaceans, large vitellogenin cDNAs have been reported that range from 7782bp in the giant freshwater prawn (*Macrobrachium rosenbergii*) (701) and 7833bp for the blue crab (*Callinectes sapdius*) (703), up to 7920bp as shown in the marine shrimp (*Penaeus semisulcatus*) (704). In *Dermancenter variabilis*, the American dog tick, the molecular weight of vitellogenin has been reported to be 462-468kDA, determined through native PAGE and gel filtration chromatography (705-707), with sequencing revealing a length of 5744bp (708). Partial synthesis and characterisation of vitellogenin has been achieved in a few tick species (705, 707, 709, 710). The first complete amino acid sequence for tick vitellogenin was found in the fat body, midgut and ovary of *Dermacentor variabilis*, alongside mRNA evidence for a potential second vitellogenin (708). In the black-legged tick, *Ixodes scapularis*, one study identified eight subunits with molecular weights ranging from 45-145kDA (711), and the clarification of three vitellogenins was provided by the Dana-Farber Cancer Institute (DFCI) gene index project (http://compbio.dfci.harvard.edu/tgi/tgipage.html). In *H. longicorns*, three complete vitellogenins have been described (*Hl*Vg-1, *Hl*Vg-2 and *Hl*Vg-3) (697). The mRNA from these Hl-Vgs were discovered in fed females where major expression sites included the fat body, ovary and haemolymph. Western blot and native PAGE showed that these vitellogenins in these sites consisted of four major polypeptides (697). At present, there is limited information regarding the number of vitellogenins present in *D. gallinae*, with just one mRNA transcript identified at currently. Further investigation would be required to understand whether a single or a family of vitellogenins copy is present.

8.1.2.1.2 Vitellogenin as a vaccine candidate

Research has shown in *D. gallinae* that host IgY is capable of entering the haemolymph following a blood meal (200). Thus, after *D. gallinae* vitellogenin is accessible to vaccine-induced antibodies following a blood meal from an immunised host (186). In a study by Bartley et al., (2015) identifying and evaluating target vaccine antigens, they demonstrated significant levels of mite mortality when using vitellogenin as an immunogen (186). They hypothesised that due to the physiological role vitellogenin plays in oocyte maturation and other functions in arthropod species (712), immunisation of hens with vitellogenins from *D. gallinae* could cause a reduction of production or viability of *D. gallinae* eggs, resulting in delayed expansion of *D. gallinae* populations (186)

Mitchell et al. (2007) reported the first successful RNAi knockdown of a vitellogenin receptor VgR in Acari, from *D. variabilis* (713). Through disabling VgR mRNAs the receptor was rendered ineffective, causing substantial amounts of vitellogenin to accumulate in the haemolymph of treated ticks instead of in the oocytes. Northern blots showed abundance of VgR mRNA in ovaries of vitellogenic females but no VgR mRNA in other female tissues or male whole-body extracts, demonstrating that VgR expression is sex- and tissue specific. Newly emerged unfed *D. variabilis* females were placed on a rabbit host, injected with 0.5 µg of double-stranded RNA into body cavity, allowed to mate with males and feed to repletion (approximately 8 days). After collection, ticks were kept 0-4 days post drop off and results showed PBS-injected control group oocytes were almost entirely brown from vitellogenin uptake two days post drop off, whilst in contract RNAi-treated oocytes had failed to progress past stage two of their development. The females ovaries were predominately white in colour and mated females failed to lay eggs (713). Another study by Boldbaater et al., (2010) demonstrated with RNA interference, that targeted three vitellogenin genes, that *Hl*-Vg double strand RNA-injected ticks achieved a lower body weight, reduced egg weight and higher mortality in engorged females following a blood meal when compared to controls. This demonstrated that these three vitellogenins in *H. longicorns* are criticial for egg development and oviposition (697). In *Amblyomma hebraeum*, injection into female ticks with 1.0 µg VgR-dsRNA showed transcript suppression but blockage of vitellogenin entering oocytes at the level of the previous studies was lacking. It has been hypothesised *A. hebraeum* might require additional unknown vitellogenin uptake factors (VUF) for successful yolk uptake but identification of these factors remains to be done (714). Immunisation of sheep with the mature form of vitellogenin, vitellin, from *R. microplus* produced a significant reduction following challenge in the number of engorged ticks, as well reductions in weight and oviposition (710).

8.1.2.2 Cathepsin D

Cathepsin D (Cat-D) is a lysosomal aspartic endopeptidase that is soluble in nature. Synthesis of Cat-D is completed in rough endoplasmic reticulum as preprocathepsin D (715). The mature form of Cat-D consists of heavy 34 kDa and light 14 kDa chains which are linked through non-covalent interactions (716-718). Generally speaking, as an enzyme group, aspartic proteases consist of two lobes which are separated due to a cleft that contains the catalytic site made up from two aspartate residues (715). Cat-D, like other aspartic proteinases (including Cathepsin E, pepsin and renin), accommodates up to eight amino-acid residues in the binding cleft. Around the cleaved bond there is a preference towards hydrophobic residues (715). Renin proteases show a high specificity towards oligopeptidic substrates, which is narrower specificity in Cat-D (719, 720). Based on Cat-D's ability to cleave a number of functional and structural peptides and

proteins, a range of physiological functions have been suggested in mammals (715). These include the activation and degradation of hormone and growth factors, metabolic degradation of intracellular proteins, brain antigen processing, processing enzyme inhibitors and activators, regulation of programmed cell death and activation of enzymatic precursors (721-730).

8.1.2.2.1 Cathepsin D in parasites

In parasites, peptidases (proteolytic enzymes, proteases) are involved in a range of adaptive functions, including moulting, coagulation, immune evasion, tissue penetration, degradation of cellular matrixes and digestion of host blood proteins (731, 732). Aspartic peptidases (AP) of the Cat-D type (APD) are a relatively small enzyme group, when compared to other peptidase families, such as serine, cysteine, metallopeptidases, that offer therapeutic potential (732). APD's use a unique mechanism of substrate binding and activation (733) which predetermines their species and non-redundant endopeptidolysis as part of key physiological processes (732).Proteolytic enzymes have essential functions and tend to be highly conserved, therefore represent a promising group of molecules to target for vaccine candidates in ectoparasites (188).

8.1.2.2.2 The role of Cathepsin D in arthropod and acari species

Cathepsin D and L proteases have both been identified in a number of mite and tick tissues (734-736) where they are involved in cleaving multiple proteins, including vitellin, albumin, gelatin and haemoglobin (735, 737-739). Experimental vaccination with tick cathepsins involved in yolk processing have yielded encouraging results [72-74].

In arthropod phylogeny, more than 20 separate evolutionary events of blood feeding have been demonstrated (740) and independent evolution of feeding strategies, digestion of blood components and modulation of host immune response has played a pivotal role in arthropod ectoparasites as specific disease vectors (732). In the Arthropoda there are two evolutionarily distinct haematophagous groups: acari and triatomine insects (741). These two groups are unique in their use of cysteine-aspartic peptidases (CA) in the processing of dietary proteins from host blood meals and their adaptation of ancestral APDs into the initial components of an intestinal multienzyme proteolytic network (741). In ticks, confirmation of a multienzyme CA based digestive system (732) was identified utilising a complex approach in *Ixodes ricinus* (741, 742) and through the recently published tick genome. It has been shown that midgut APD endopeptidase activity initially cleaves the host haemoglobin, with high turnover rate, into large fragments (742), but does not appear to exhibit the same function in albuminolysis (743). A combination of biochemical and genetic analyses of the APD proteolytic network in *I. ricinus* indicated the presence of four cysteine peptidase paralogues (Cathepsin B, C, L and legumain)

in combination with Cat-D, operating together for haemoglobinolysis (744). Gene expression revealed that increases in total haemoglobinolysis were matched to the activity profile of Cat B, C, D and legumain (745). High expression of Cat-D has been demonstrated in the midgut of *H. longicornis* after ingestion of a blood meal, as well as expression in the salivary glands. In *D. gallinae*, Bartley et al., (2015) identified a 383 amino acid protein (Dg-CatD-1) that shows homology cathepsin D lysosomal aspartyl proteinases (188). More recently, Price et al., demonstrated that purified refolded recombinant cathepsin D acted as an active aspartyl proteinase, digesting haemoglobin with a pH optimum of pH 4 (201).

8.1.2.3 Tropomyosin and Paramyosin

Tropomyosin and paramyosin are microfilament proteins which are involved in facilitating the interaction between actin with troponin and myosin throughout filament assembly and contraction (192, 746). Whilst tropomyosin is found ubiquitously in eukaryotic cells with multiple tissue specific forms identified (747, 748), paramyosin is known to be specific to invertebrates, acting as a component of the thick filament of muscles (749). Previous research in ecto- and endoparasites has demonstrated promising results for paramyosin and tropomyosin as vaccine candidates. Examples include the trematode *Schistosoma japonica*, the filarioid nematode *Acanthocheilonema viteae* and the tick *H. longicornis* (750-752). At present, the mechanisms that result in the mortality inducing effects of paramyosin and tropomyosin vaccination are not fully understood (192). It is hypothesised that antibody binding to *D. gallinae* tissues could cause a disruption in function or formation of the cytoskeleton (753).

8.1.2.3.1 Tropomyosin

Tropomyosin is an actin-binding, allergenic protein (200). It has been shown to be recognised by IgE in sera from 5.6-81% of humans with a house dust mite allergy (754, 755). This variability, at least partially, may result from the sera source as well as prior sensitisation of individuals to tropomyosins from other species cross-reacting (756). Research into using tropomyosin as a vaccine candidate against parasites has shown that immunisation of jirds with a tropomyosin like molecule from the nematode *A. viteae* demonstrated a >60% reduction in adult worm burden after challenge, as well as a reduction in circulating microfilariae of up to 93% (750). Immunisation of rabbits with a recombinant tropomyosin from *H. longicorns* resulted in significant reductions ($P<0.05$) in tick engorgement weights (19.4%), egg mass (14.7%), egg hatching rate (100%) and oviposition (49.5%) (751).

In acari species, characterisation of tropomyosin has been undertaken for a few mite species. In *D. gallinae* one gene with two isoforms of tropomyosin have been identified, comparison with

Metaseiulus occindetalis, demonstrated 94% amino acid similarity with *D. gallinae* isoform one and 97% amino acid similarity to *D. gallinae* isoform two, whilst comparison with the house dust mites *Dermatophagoides pteronyssinus* and *D. farinae* (85% and 86% identity respectively) (754, 755) and *Psoroptes ovis*, the sheep scabies mite, showed 86% identity (746). One study demonstrated through Western blotting using anti-HDM tropomyosin sera that tropomyosin is present in extracts of *P. ovis,* which provokes IgE and IgG responses in the sheep host during infestation (746). Tick orthologues for tropomyosin have also been identified in *R. microplus* (89% similarity to *D. gallinae*) and *H. longicorns* (88% similarity to *D. gallinae*) (200).

8.1.2.3.2 Paramyosin

Work by Wright et al., (2016) demonstrated through an *in vitro* feeding assay that antibodies raised against a recombinant version of paramyosin significantly increased mortality of *D. gallinae* by 23% (192). The results from their study demonstrated that at the time, paramyosin was one of the best performing recombinant vaccine candidates that had been tested regarding *D. gallinae* mortality. Previous research has supported the use of paramyosin as a vaccine candidate for the use in parasite control. Immunisation of mice with the native form of paramyosin from *Schistosoma japonica* showed a reduction in adult worm burden (up to 86%) following challenge with the parasite (752). Another study using a recombinant cocktail vaccine containing fragments of *S. japonica* fragments in BALB/c mice also demonstrated a reduction in worm burden (up to 40%) and a reduction in liver-stage eggs (up to 78%) following parasite challenge (757).

8.2 AIMS AND HYPOTHESES

8.2.1 Aim of the study

The main aim of this study was a gene-specific genetic analysis focusing on assessment of existing genetic diversity and signatures of selection at loci encoding prominent anti-*D. gallinae* vaccine candidates

8.2.2 Hypotheses

Hypothesis 10: Genetic diversity is expected to be minimal at loci encoding candidate vaccine antigens, resulting from (i) lack of exposure to immune selection, and (ii) putative conserved protein function

Hypothesis 11: Vaccine candidates will be present signatures of strong purifying selection, showing significant deviation from neutrality

8.3 METHODOLOGY

8.3.1 Sample selection

8.3.1.1 Sanger sequencing

A combination of single and pooled mite gDNA extracts were selected for amplification of candidate vaccine genes (Table 42). Twelve UK farms, including all four countries, were represented by individual gDNA. A further seven pooled gDNA samples were included from farms located elsewhere in Europe. Additionally, samples from three of the 12 UK farms were selected for amplification of cDNA (Table 42).

Extract type	Sample	Production system
Single mite gDNA extracts	UK1	Free-range
	UK2	Free-range
	UK3	Free-range
	UK6	Free-range
	UK7	Free-range
	UK8	Intensive
	UK9	Free-range
	UK11	Free-range
	UK12	Free-range
	UK13	Intensive
	UK20	Free-range
	UK23	Free-range
Pooled gDNA extracts	BEL5	Intensive
	DEN2	-
	FRA3	Intensive
	NET7	Intensive
	POR1	Intensive
	ROM6	Backyard
	SPA4	Intensive
Pooled cDNA extracts	UK1	Free-range
	UK6	Free-range
	UK9	Free-range

Table 42: *Type of extractions conducted with associated sample and production systems used for amplification of vaccine candidates relating to* D. gallinae

8.3.1.2 SNP genotyping with Multi-plex

Samples selected for SNP genotyping are outlined in Chapter 6 (Section 2.3.1 Sample selection). Briefly, pooled *D. gallinae*, including 50-80 mites per pool, were used from farms across the UK and Europe.

8.3.2 DNA extraction

DNA extraction was completed using the protocol outlined in General Methodology 2.3.

8.3.3 RNA Extraction

Two methods of RNA isolation were used: a commercial RNeasy RNA kit (8.3.3.1) and Trizol and Chloroform extraction (8.3.3.2). This was predominantly due to limited laboratory access during COVID regulations and delays to supplies arriving to the RVC at the time of the work being conducted.

8.3.3.1 RNeasy RNA kit

RNA extraction was completed from pooled mites (~50-100 mites) following the manufacturer's protocol for a Qiagen RNeasy Mini kit (Qiagen, QmBH) with homogenisation using a traditional

pestle and mortar in 600µl of buffer RLT, as specified for animal tissues stabilised in RNA*later* or difficult-to-lyse tissues.

8.3.3.2 Trizol and Chloroform Extraction

Approximately 50-100 mites were homogenised in 1ml of Trizol® (Thermo Fisher Scientific, Waltham, Massachusetts, USA) using a traditional pestle and mortar and incubated for five minutes at room temperature to permit complete dissociation of the nucleoprotein complex. 200µl of chloroform (Fisher Scientific, Hampton, New Hampshire, US) was added to each sample and mixed vigorously for 10 seconds before being incubated for three minutes at room temperature. Samples were centrifuged at 12,000 x g for 15 minutes at 4°C. After centrifugation, 400µl of the aqueous upper phase was transferred to a new 2µl Eppendorf tube and 800µl of cold isopropanol was added (isopropanol was stored at ~-20°C for one hour prior to RNA extraction). Tubes were mixed through inversion and then incubated on ice for two minutes before centrifugation at maximum speed (10,380 g) for 25 minutes at 4°C. The supernatant was removed, and tubes were centrifuged for a further 30 seconds and the remaining supernatant removed. The pellet was washed with 600µl of cold 70% (v/v) ethanol (stored at ~-20°C for one hour prior to RNA extraction) and the tube flicked before spinning for 10 minutes at maximum RCF (g) for 10 minutes. Ethanol was removed from the tube and the wash step repeated with 400µl of 70% ethanol. After spinning, all liquid was removed, and the pellet air dried for up to ten minutes. RNA was dissolved in 50µl of RNAse free water. Quantification was performed on a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) following the standard manufacturer's protocol. Samples were stored at -80°C.

8.3.4 TurboDNAse treatment

TurboDNAse (Invitrogen, Waltham, Massachusetts, USA) was used for degradation of any DNA still present in samples to improve RNA integrity. 5µl Turbo 10X buffer, 5µl RNA sample, 1µl of TurboDNAse and 39µl DNase and RNase free water (Invitrogen, Paisley, UK) were combined and placed in the thermocycler for 30 minutes at 37°C. To prevent TurboDNAse from interfering in downstream applications it was inactivated using Ethylediaminetetraacetic acid (EDTA). 1.5µl of 0.5mM EDTA was added to 50µl reaction mixture and put in the thermocycler for 10 minutes at 75°C. Products were stored at -80°C.

8.3.5 Reverse Transcription PCR (RT-PCR)

Two methods of RT-PCR were completed as stocks of Superscript IV became limiting during the work, when iScript™ cDNA synthesis kit was used.

8.3.5.1 Superscript IV

To anneal primer to template RNA, 1µl of 50µM random hexamers, 1µl of 10µM dNTP (10µM each), 10µl of template RNA and 1µl of nuclease free H_2O were combined together in a 1.5ml Eppendorf tube and mixed via brief vortex followed by centrifugation. Samples were incubated at 65°C for five minutes and then incubated on ice for five minutes. To prepare the RT-PCR mixture, 4µl of 5 x SSIV buffer, 1µl of 100mM DTT, 1µl of RNAseOUT™ recombinant RNA inhibitor (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and 1µl of Superscript® IV Reverse Transcriptase (200µ/ml) (Thermo Fisher Scientific, Waltham, Massachusetts, USA) were combined, mixed by vortex and briefly centrifuged. The RT-PCR mixture and annealed RNA mixture were combined and incubated at 23°C for 10 minutes followed by 55°C for 10 minutes and finally at 80°C for 10 minutes. Products were stored at -20°C.

8.3.5.2 iScript cDNA synthesis kit

For cDNA synthesis with iScript™ (Bio-Rad, Hercules, California, USA), 4µl of 5 iScript Reaction Mix, 1µl of iScript Reverse transcriptase, 10µl of RNA template and 5µl of nuclease free water were combined together and incubated in a thermocycler for five minutes at 25°C, 20 minutes at 46°C, one minute at 95°C and held at 4°C. Products were stored at -20°C.

8.3.6 Primers

Primers were designed as described in General methodology 2.6 with the following primer pairs utilised (Table 43-44).

8.3.6.1 Cathepsin D

Primers for Cat-D were designed using the mRNA sequence produced by Bartley et al., (2012) (188). GenBank Accession Number: HE565350.1. A total of 12 primers were designed (Table 43).

Primer name	Primer sequence	Length	TM (°C)	GC%	Secondary structure	Primer dimer
CatD_For1	CGATCTCATCAGGGTGCCTCTG	22	70.3	59.10	Weak	No
CatD_Rev1	GCTTTGCCGATGCCGCATAA	20	72.0	55.00	Weak	No
CatD_For2	TACGGTCACCAAGCAGACGTT	21	66.7	52.40	Weak	No
CatD_Rev2	CGAAGAGTGACGTACCCGTGTC	22	67.9	59.10	Weak	No
CatD_For1B	CAGGGTGCCTCTGAAAAAGAT	21	64.30	47.62	Weak	No
CatD_Rev1B	GCCTACGCGATTGTTTTCAC	20	64.40	50.00	None	No
CatD_For2B	TGGCGAGATCACAGAGGAGT	20	65.10	55.00	None	No
CatD_Rev2B	TCTAACCCGCGAAAGCTACTGG	22	68.20	54.55	None	No
CatD_For4	TCGCGGCCCTGTGCAC	16	71.70	75.00	None	No
CatD_Rev4	TTGGCATGGGCTATCCGGA	20	70.80	57.89	None	No
CatD_For5	GGCTCGTCGGATCTTTGGGT	20	69.60	60.00	None	No

Table 43: *Primer sequences, length, melting temperature, GC%, presence of secondary structure or primer dimer and corresponding region of mRNA sequence for D. gallinae Cathepsin D*

8.3.6.2 Vitellogenin

Primers for vitellogenin were designed using the mRNA sequence produced by Bartley et al., (2015) (186) (Table 44). GenBank Accession Number: KR697567.1.

Primer name	Primer sequence	Length	TM (°C)	GC%	Secondary structure	Primer dimer
Vit_For1	ATGAGGTTCTTCGTTCTCCCT	21	62.9	47.6	Weak	No
Vit_Rev1	CAGAAGAGCATAGACGGGAATG	22	64.7	50.0	Very weak	No
Vit_For2	AACCCGCTACACTTGCTGT	19	62.4	52.6	Very weak	No
Vit_Rev2	CCGTTGATTGTCTTCTTGTGTT	22	62.8	40.9	None	No
Vit_For3	CGCACGAGGAAGATGAGCTT	20	66.8	55.0	Very weak	No
Vit_Rev3	ACAATCGGGTTAACCTCTTTCG	22	64.9	45.5	None	No
Vit_For4	CGCCACCGGTAAAGCTACAG	20	66.6	60.0	Very weak	No
Vit_Rev4	TCTCGCGTTTCAGACCCATA	20	65.6	50.0	Very weak	No
Vit_For5	CGGTATTCGCGGAGTTTGGA	20	69.1	55.0	None	No
Vit_Rev5	GAGAATGCCTTCGCTGCTTACA	22	67.5	50.0	Very weak	No
Vit_For6	CTTTCCACGTCGAGCCCTCTAC	22	68.1	59.1	None	No
Vit_Rev6	TGACGGCGATTTTCAATTGG	20	67.8	45.0	Very weak	No
Vit_For7	TCCGCTACGACCCTACCTTCT	21	65.9	57.1	None	No
Vit_Rev7	TGAACGTTGGAAGTGTCTACCA	22	64.4	45.5	Weak	No
Vit_For8	CGATCGCTACATCACTGTTGTG	22	65.5	50.0	Weak	No
Vit_Rev8	TTAATCGTGACGTTCGCACC	20	65.9	50.0	Weak	No

Table 44*: Primer sequences, length, melting temperature, GC%, presence of secondary structure or primer dimer and corresponding region of mRNA for* D. gallinae *vitellogenin*

8.3.7 PCR and agarose gel electrophoresis

PCR and agarose gel electrophoresis were carried out following the protocols outlined in General Methodology 2.7 and 2.9, respectively.

8.3.8 PCR purification and Sanger sequencing

PCR purification and Sanger sequencing were carried out following the protocols outlined in General Methodology 2.10 and 2.11, respectively.

8.3.9 Nucleotide analysis

Population diversity indices including haplotype diversity, number of haplotypes, nucleotide diversity (π) and the average number of nucleotide differences (k) were calculated for each PCR amplicon. These tests were all done using DnaSP version 6.12.03 (593). Nucleotide diversity was defined as the average number of nucleotide differences per site among DNA sequences by pairwise comparison, whilst haplotype diversity (also referred to as gene diversity) was representative of the probability that two randomly sampled alleles will differ (345).

8.3.10 Neutrality tests

8.3.10.1 Tajima's D

Analysis of Tajima's D was completed using DnaSP version 6.12.03 (593). The same theory and assumptions as outlined in chapter three (Genetic Markers for *Dermanyssus gallinae*), section 2.3.4.2.1. (Tajima's D) were followed.

8.3.10.2 Fu and Li's D and F test

Analysis of Fu and Li's D and F was completed using DnaSP version 6.12.03 (593). The same theory and assumptions as outlined in chapter three (Genetic Markers for *Dermanyssus gallinae*), section 2.3.4.2.2. (Fu and Li's D and F test) were followed.

8.4 RESULTS

In order to achieve the aim of this study, one major objective was the amplification of loci encoding all four vaccine candidates from cDNA and/or gDNA for Sanger sequencing. Due to time constraints, complications related to PCR of vaccine candidates and COVID-19 limiting laboratory access, it was not possible to assembly full loci for each target. Thus, many analyses were undertaken using amplicons representing fragments of each locus.

8.4.1 Cathepsin D cDNA

Amplification from cDNA using six pooled samples representing three UK farms (UK1, UK6 and UK9) revealed 24 SNP mutations in comparison to the reference sequence (Table 45). Two pooled samples were sequenced in both forward and reverse directions (UK6.1 and UK6.2), No insertions or deletions were observed (Table 45).

Reference position	Reference Cat-D residue (HE565350.1)	Mutation	No. of populations with reference	No. of populations with SNP
241	A	G	4	2
268	T	G	5	1
342	A	T	5	1
344	C	T	5	1
357	G	A	4	2
373	G	A/C	3	3
447	C	T	4	2
456	A	G	2	4
475	G	A	3	3
509	A	A	3	3
552	C	T	0	6
600	G	A	1	5
652	G	C	5	1
724	A	G	5	1
725	G	C	5	1
749	A	C	5	1
786	G	A	5	1
809	A	G	5	1
823	C	A	5	1
863	G	A	5	1
867	G	T	5	1
987	C	A	4	2
997	T	A	2	4
1020	C	A	5	1

Table 45: *Polymorphism identified across Cathepsin-D from* D. gallinae *cDNA, consensus residue, reference (Cathepsin D mRNA sequence) residue, mutation, no. of populations with reference residue and no. of individuals with SNP outlined*

8.4.1.1 Nucleotide, diversity, haplotype diversity and neutrality tests for Cathepsin D cDNA

Analysis of six pooled *D. gallinae* samples revealed high haplotype diversity (1.000), attributed to six haplotypes for each population analysed (Table 46). No significant deviations were seen for Tajima's D and Fu and Li's D and F statistic.

Category	Value	Significance
Nucleotide diversity	0.01132	-
Haplotype diversity	1.000	-
No. of haplotypes	6	-
Average no. of nucleotide differences	9.467	-
Tajimas's D	-0.10958	Not significant
Fu and Li's D	0.08276	Not significant
Fu and Li's F	0.04546	Not significant

Table 46: *Summary of nucleotide diversity, haplotype diversity and neutrality tests for cDNA amplified for* D. gallinae *Cathepsin D*

8.4.1.2 Cathepsin D: gDNA

Amplification of all primer pairs failed for gDNA except for CatD_For2 and CatD_Rev2. Amplification using these primers generated a product of ~650bp, larger than the 346bp expected from the cDNA sequence. Amplification from individual *D. gallinae* DNA extracts failed, with amplicons from just four pooled *D. gallinae* samples (BEL5, FRA3, NET7 and ROM6 (Table 42) sequenced. Sequence analysis revealed alignment of the first 138bp of gDNA sequence to the cDNA, corresponding to nucleotide positions 531-668 of the reference Cat-D sequence. Eight SNPs were identified (Table 47). The remaining sequence failed to align to the Cat-D cDNA reference, but all four sequences aligned to each other, indicating the presence of an intron in *D. gallinae* Cat-D beginning at nucleotide position 669.

Reference position	Consensus residue	Reference Cat-D residue	Mutation	No. of populations with reference	No. of populations with SNP
552	T	C	T	1	3
572	A	A	C	3	1
578	T	T	A	3	1
600	A	G	A	1	3
650	G	G	A	3	1
655	C	C	A	3	1
659	T	T	C	3	1
655	G	G	A	3	1

Table 47: *Nucleotide polymorphisms found in gDNA from four pooled mite samples (ROM6, BEL5, FRA3 and NET7) in comparison to the reference* D. gallinae *CAT-D sequence*

8.4.2 Vitellogenin

A total of 91 polymorphic sites were identified using gDNA extracted from individual *D. gallinae* across a 3406bp sequence assembly, represented by six partially overlapping fragments amplified using different primer pairs (Table 48). Five of the polymorphic sites were insertions/deletions and the remaining 86 were SNPs. (Table 49). Five sites were tri-allelic, all in fragment four, and the remaining 86 were biallelic (Table 49).

Fragment number	Primers used	Amplicon size	No. of individual *D. gallinae* gDNA extracts	No. of pooled *D. gallinae* gDNA extracts
2	Vit_For2, Vit_Rev2	974bp	14	0
3	Vit_For3, Vit_Rev3	629bp	11	0
4	Vit_For4, Vit_Rev3	44bp	74	0
6	Vit_For6, Vit_Rev5	514bp	24	0
7	Vit_For7, Vit_Rev7	882bp	6	2
8	Vit_For8, Vit_Rev8	806bp	11	4

Table 48: *Fragment number, corresponding primers used, resulting amplicon size and no. of individual gDNA* D. gallinae *and pooled* D. gallinae *extracts for six fragments of vitellogenin amplified for Sanger sequencing*

Primer pair/Fragment	Reference position	Consensus residue	Reference Vitellogenin residue	Mutation	No. of individuals with reference	No. of individuals with SNP
Vit_For2, Vit_Rev2 (Fragment two)	570	C	C	A	11	3
	633	G	C	G	6	8
	668	A	A	C	13	1
	759	C	C	T	13	1
	804	G	A	G	6	8
	834	C	C	T	13	1
	857	T	T	C	9	5
	921	T	T	C	13	1
	1119	G	G	A	12	2
	1135	G	G	1	13	1
	1191	G	T	G	7	7
	1330	G	G	A	13	1
	1344	T	T	C	12	2
	1399	A	A	G	13	1
	1447	C	C	T	12	2
	1460	T	T	C	13	1
	1469	A	A	T	13	1
	1485	T	T	A	13	1
Vit_For3, Vit_Rev3 (Fragment 3)	1113	T	T	A	9	2
	1119	G	G	A	9	2
	1135	A	G	A	4	7
	1139^1140	A	-	A	5	6

	1140	C	C	A	7	4
	1180	A	A	C	7	4
	1191	G	T	G	5	4
	1203	T	G	T	6	5
	1208^1209	-	-	G	9	2
	1209	G	T	G	3	8
	1314	T	C	T	3	8
	1344	C	T	C	0	11
	1350	T	T	C	9	2
	1377	T	C	T	1	10
	1399	A	A	G	6	5
	1414	A	G	A	4	7
	1419	C	T	C	3	8
	1447	T	C	T	3	8
	1460	T	T	C	10	1
	1486	A	G	A	3	8
	1533	C	C	A	9	2
	1553	A	C	A	3	8
	1590	G	G	A	10	1
	1656	G	G	A	10	1
	1683	G	G	T	10	1
	1690	G	G	A	10	1
Vit_For4, Vit_Rev3(Fragment 4)	1581	T	G	A/T	29	10/35
	1582	T	T	A/G	66	4/4
	1583	G	T	A/G	35	4/35
	1584	G	G	C	73	1
	1585	C	G	A/T	35	1/38
	1586	T	C	A/T	35	1/38
	1587	A	T	A	36	36
	1588	C	C	A	72	2
	1703^1704	-	-	A	73	1
	1848	G	G	A	73	1
	1956	G	G	A	54	20
	1962	C	C	T	0	46
	1972	G	G	A	67	7
	1974	A	A	C	46	29
	1998	A	G	A/C	26	48/1
	2015	A	A	C	74	1
	2019	G	G	C	66	9
	2020	A	A	G	66	9
	2021	T	T	A	66	9
	2023	G	G	T	66	9
Vit_For6, Vit_Rev5 (Fragment 6)	2877	A	A	G	12	12
	3014	A	A	G	17	7
	3060	C	C	T	14	10
	3118	A	G	G	24	24
	3124	G	G	A	13	11
	3165	T	T	C	15	9
	3194	C	T	C	21	3
	3221	A	A	G	22	2
	3258	C	C	T	12	12

	3264	G	G	A	21	3
Vit_For7, Vit_Rev7 (Fragment 7)	3483	T	T	C	7	1
	3483^3484	A	A	-	5	3
	3490	G	G	T	7	1
	3514	A	A	T	7	1
	3582	G	G	T	7	1
	3671	A	A	G	5	3
	3672	G	A	G	1	7
	3688	A	A	G	5	3
	4219	G	G	A	5	3
	4235	A	A	T	7	1
	4237	C	G	C	2	6
	4241^4242	-	-	T	5	3
	4245	G	G	T	7	1
	4307	G	G	C	7	1
	4310	A	A	T	7	1
	4317	A	A	T	4	4
Vit_For8, Vit_Rev8 (Fragment 8)	4722	C	C	T	13	2
	4782	G	A	G	3	12
	4956	C	G	G	3	12
	5013	A	A	G	13	2
	5091	C	C	T	14	1
	5130	G	G	A	9	6
	5316	T	C	T	5	10
	5526	A	A	C	9	5

Table 49: *Polymorphism identified across vitellogenin according to primer pair, consensus reside, reference (vitellogenin mRNA sequence) residue, mutation, no. of individuals with reference residue and no. of individuals with SNP/indel outlined. ^ indicating an insertion between two nucleotide positions. – indicating no nucleotide present at that site and a / indicating two alternative alleles present at tri-allelic sites*

8.4.2.1 *Nucleotide diversity for Vitellogenin*

Nucleotide diversity for vitellogenin amplicon alignments ranged from 0.00479 to 0.01337, with an average nucleotide diversity across the full vitellogenin dataset of 0.04705 (Table 50). The highest nucleotide diversity was observed in fragment four (corresponding to base pairs 1581-2023 of the reference coding sequence) and the lowest was observed in fragment 2 (corresponding to base pairs 514-1514). The nucleotide diversity of fragment four can be attributed to four clusters of polymorphisms in sliding window analysis (Figure 53). The average number of nucleotide differences across all fragments was 4.9767, ranging from 2.819 to 7.745 (Table 50).

Corresponding region of vitellogenin coding sequence (bp)	Fragment no.	Sequence length	Sequence conservation	No of samples	Sample locations	Nucleotide diversity	Average no. of nucleotide differences	Haplotype diversity	No. of haplotypes
514-1514	2	974	0.982	14	UK	0.00479	4.67033	0.934	11
1093-1719	3	629	0.967	11	UK	0.01239	7.74545	1.00	11
1581-2023	4	444	0.957	74	UK	0.01337	5.92225	0.986	51
2861-3374	6	514	0.982	13	UK	0.00748	3.846	1.00	13
3459-4338	7	882	0.984	8	UK, ROM, PORT	0.00552	4.857	1.00	8
4721-5526	8	806	0.990	15	UK, DEN, FRA, PORT, ROM, SPA	0.00350	2.819	0.981	13

Table 50: *Nucleotide and haplotype diversity of* D. gallinae *for vitellogenin, primer pair, sequence length, sequence conservation, no. of samples included in analysis, nucleotide diversity, average number of nucleotide differences, haplotype diversity and number of haplotypes*

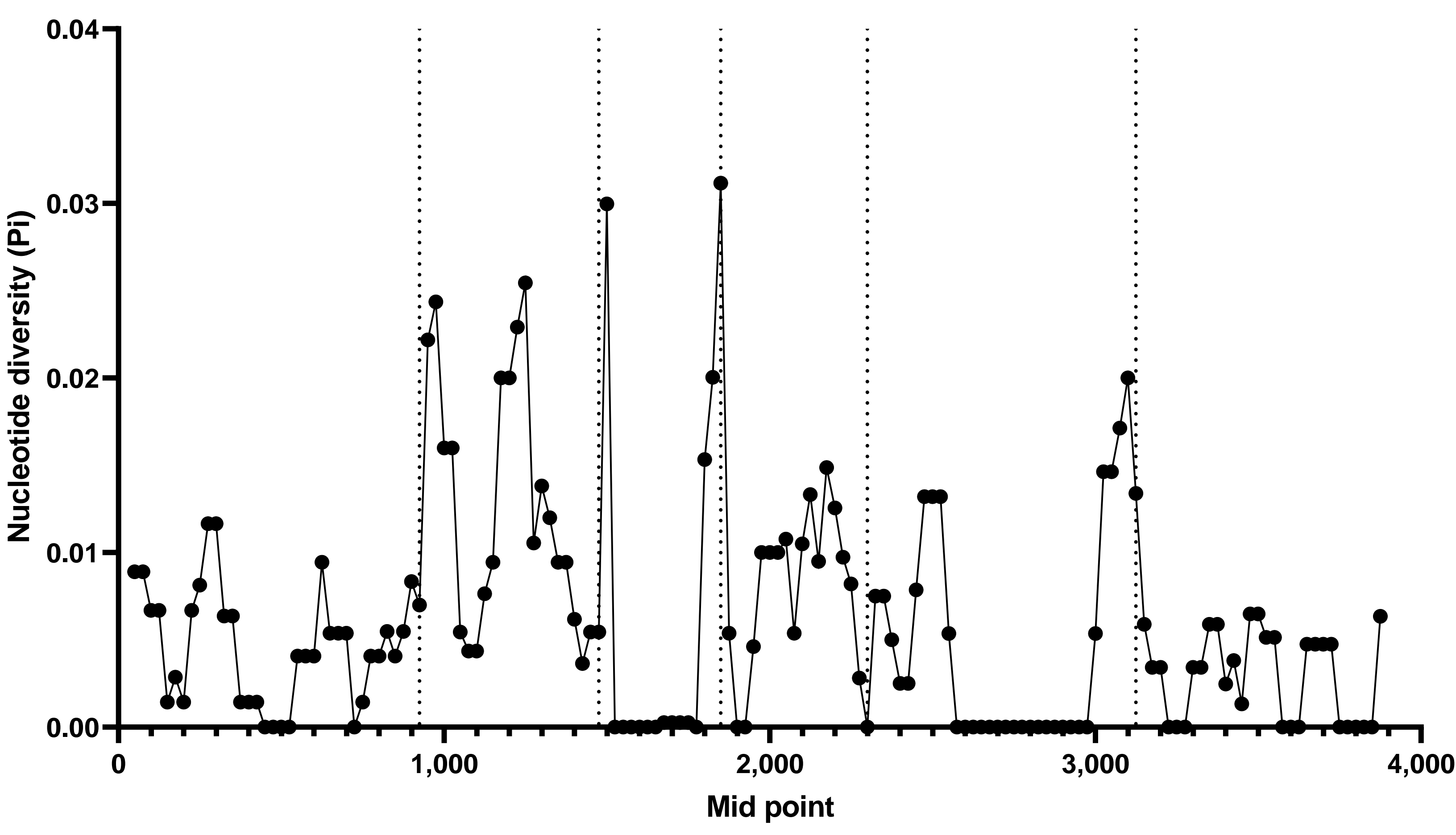


Figure 53: *Sliding window analysis of nucleotide diversity across vitellogenin for* D. gallinae. A sliding window size of 100 was used with a step size of 25. *Dashed lines indicating the end and start of consecutive primer pairs. Primer pairs put in chronological order (2,3,4,6,7 and 8)*

8.4.2.2 Haplotype diversity for Vitellogenin

Haplotype diversity ranged from 0.934 to 1, with three fragments scoring 1 as all samples represented an individual haplotype (Table 51). Fragments two, six and seven showed a unique haplotype for each individual or pooled sample analysed, fragments two and eight showed one shared haplotype across populations and fragment seven showed eleven shared haplotypes, shared by a maximum of five samples (Table 51).

Fragment no.	No. of sequences	Total no. of haplotypes	Haplotype diversity	Haplotypes found in more than one sample				No. of unique haplotypes
				Average	Min	Max	Max no. of samples	
2	14	11	0.934	1	1	1	4	11
3	11	11	1.00	0	0	0	0	11
4	74	51	0.986	3	2	4	5	51
6	13	13	1.00	0	0	0	0	13
7	8	8	1.00	0	0	0	0	8
8	15	13	0.981	1	1	1	2	13

Table 51: *Haplotype diversity for vitellogenin fragments amplified from* D. gallinae*, with summary of haplotypes identified in one than one sample*

8.4.2.3 Conserved domains

Three putative conserved domains were identified during use of BLASTP (758, 759) for comparison of amino acid composition from the Conserved Domains Database (CDD) (760-762) using the Subfamily Protein Architecture Labelling Engine (SPARCLE) tool (763). A lipoprotein N-terminal domain (LPD) at amino acid positions 31-730 (93-2130bp), a von Willebrand factor type D domain at amino acid positions 1487-1655 (4461-4965bp) and a domain of unknown function (DUF1943) at amino acid positions 763-987 (2208-2961bp) (Figure 54).

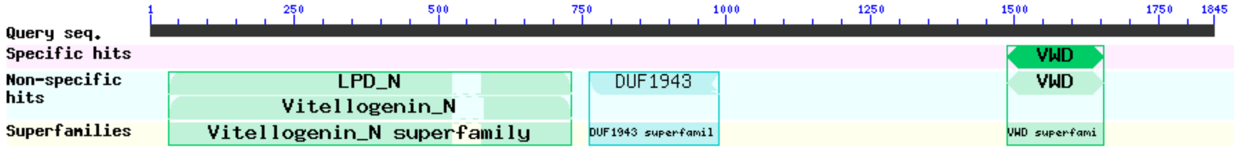


Figure 54: *Putative conserved domains in* D. gallinae *amino acid sequence as predicted by BLASTP (758, 759) and the CDD database and SPARCLE with interval locations*

8.4.2.4 Neutrality tests for Vitellogenin

Results for Tajima's D and Fu and Li's F and D test all revealed no significant deviation from zero, with sliding window analysis, revealing similar patterns between each test across all primer pairs (Figure 55).

8.4.2.4.1 Tajima's D

Tajima D values ranged from -0.72489 to 1.28487 for different amplicons from the vitellogenin locus, with an average of 0.23821 (Table 52). No alignments significantly deviated from zero, indicating neutrality in the vitellogenin gene.

Corresponding region of vitellogenin	Primer pair	Number of samples	Tajima's D value	Statistical significance
514-1514	2	14	-0.72489	NS P<0.10
1093-1719	3	11	0.36494	NS P<0.10
1581-2023	4	74	0.47889	NS P<0.10
2861-3374	6	13	1.28487	NS P<0.10
3459-4338	7	8	-0.51141	NS P<0.10
4721-5526	8	15	0.53687	NS P<0.10

Table 52: *Tajima's D values calculated for each primer pair covering vitellogenin for* D. gallinae *isolates, corresponding region of vitellogenin covered, number of samples involved in analysis and statistical significance*

8.4.2.4.2 Fu and Li's D and F test

No significant deviation was demonstrated for Fu and Li's D and F statistics for all alignments, indicating neutrality in the vitellogenin gene (Table 53).

Corresponding region of vitellogenin	Primer pair	Analysis using only biallelic positions			Results from DnaSP V5			
		No. of segregating sites	D	F	D statistic	P	F statistic	P
514-1514	2	18	-0.98527	-0.95516	-0.98527	NS	-1.04853	NS
1093-1719	3	21	0.37270	0.38204	0.37270	NS	0.42015	NS
1581-2023	4	13	-0.16607	0.03398	-0.28067	NS	0.00089	NS
2861-3374	6	9	0.95617	1.08216	0.95617	NS	1.18813	NS
3459-4338	7	14	-0.71032	-0.67334	-0.71032	NS	-0.73721	NS
4721-5526	8	8	0.84472	0.79634	0.84472	NS	0.87290	NS

Table 53: *Fu and Li's D and F values calculated for sequence generated from each primer pair for* D. gallinae *isolates for vitellogenin, no. of segregating sites, D and F values provided by biallelic positions only and D and F statistic with associated P value from DnaSP version 5 also provided*

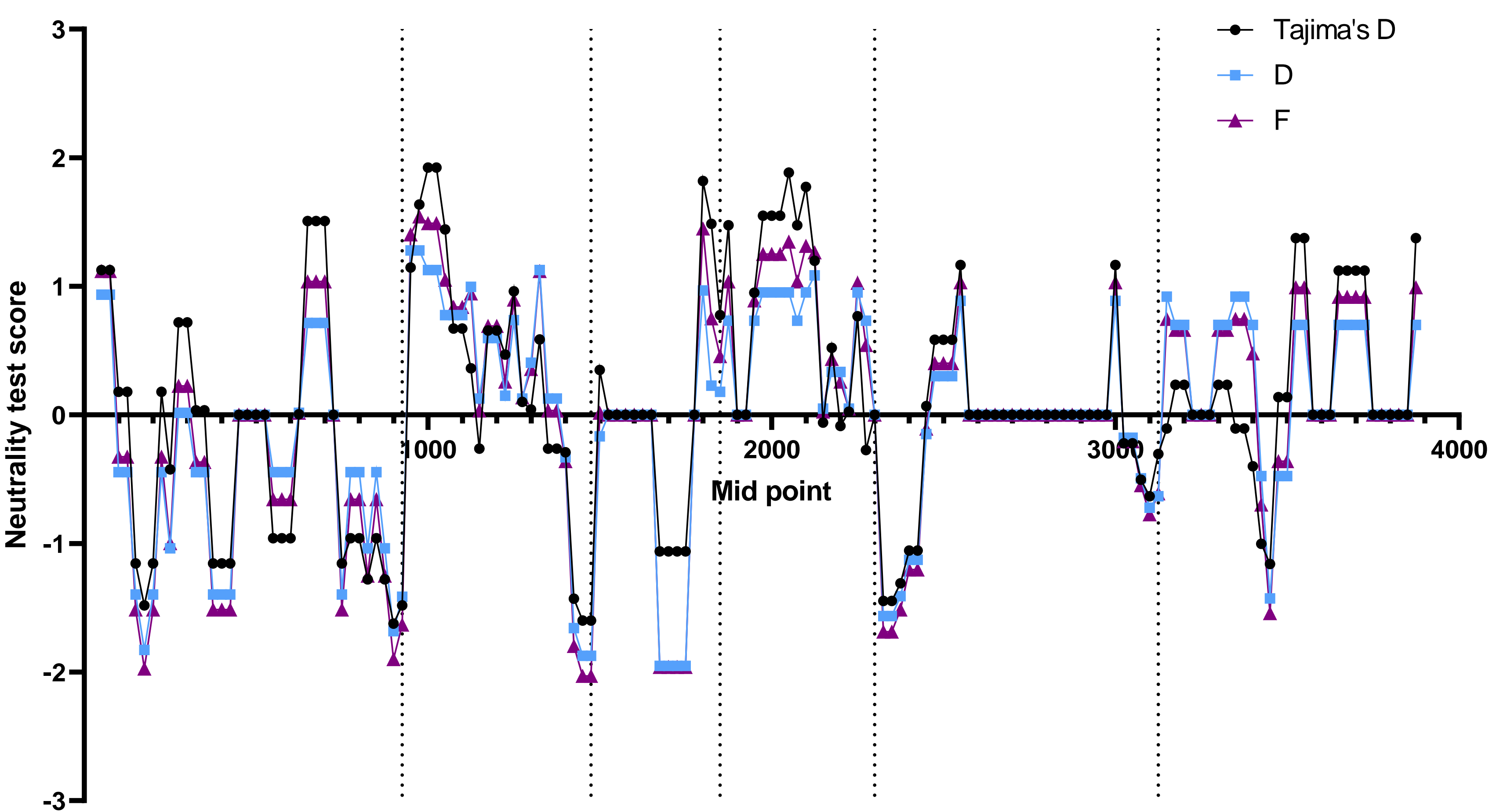


***Figure 55**: Sliding window comparison of Fu and Li's F and D test and Tajima's D test for all primer pairs across vitellogenin for all D. gallinae. Sliding window computed on DnaSP, with a sliding window length of 100 sites and step size 25. Dashed lines indicating where primer pair alignments finish and begin, in chronological order (2, 3, 4, 6, 7 and 8)*

8.4.1 SNP genotyping of vaccine candidates using the Mid Plex sequencing assay

A total of 17 SNPs were analysed from 76 pooled *D. gallinae* samples, covering the UK and 12 additional countries from Europe (Table 54). The dominant allele in each population sample was used in this analysis (see Chapter 5, Section 2.3.7.1 Conversion of heterozygous allele calls to dominant allele for methodology). Eleven SNPs related to vitellogenin, one related to Cathepsin-D and one related to Paramyosin were genotyped. Amplification of tropomyosin failed to produce analysable fragments so have been excluded. Six samples showed reference alleles for all 17 SNPs across the four loci (four farms from the UK, one farm from Portugal and one farm from the Netherlands). Alternative alleles were detected at 15 of the marker sites, with no alternative alleles observed for one vitellogenin marker (Table 54). From 11 SNP markers located in the vitellogenin gene, seven had a frequency of less than 5%, with the most frequently occurring SNP being present in 31.6% of the population.

8.4.1.1 Nucleotide diversity in amplified vaccine candidate fragments

Raw sequencing reads for each sample were provided by Eurofins and viewed on Integrative Genomics Viewer (IGV) (764) to identify any additional polymorphisms present in the regions amplified for each vaccine candidate. No additional mutations were observed in the cathepsin D or paramyosin sequence fragments for any of the samples analysed (Table 54).

A total of 29 SNPs were discovered across two fragments from the vitellogenin locus amplified (287bp and 403bp, respectively), with 11 of these coded by SNP genotyping. All 29 SNPs displayed heterozygosity at all positions for all samples analysed, demonstrating reference and alternative alleles present at these positions in all *D. gallinae* populations analysed (Table 54).

SNP (REF/ALT)	No. of reference	No. of alternative	N/A	% of total population with alternative allele
		Cathepsin D		
G/A	44	32	0	42.1
		Paramyosin		
C/A	61	15	0	19.7
		Tropomyosin		
G/A	76	0	0	0
T/C	70	6	0	7.9
G/C	46	30	0	39.5
A/C	35	41	0	53.9
		Vitellogenin		
A/C	73	0	3	0
C/T	69	3	4	3.9
C/T	59	13	4	17.1
C/G	50	22	4	28.9
C/T	50	22	4	28.9
T/C	48	24	4	31.6
T/C	73	1	2	1.3
A/G	73	2	1	2.6
T/A	73	2	1	2.6
C/T	73	2	1	2.6
T/C	73	2	1	2.6

Table 54: *Mid-plex SNP genotyping results for 17 SNPs across four loci (Cat-D, paramyosin, tropomyosin and vitellogenin) of 76 pooled* D. gallinae *samples with number of reference, alternative and NA (failed to amplify) samples and the percentage of the total population (defined here as all 76 populations studied).*

8.4.1.2 Non-synonymous vs synonymous mutations

Fragments of Vitellogenin amplified through Sanger sequencing demonstrated a dN/dS ratio ranging from 0.14 to 0.67 and 2-10 haplotypes, whilst Cathepsin D demonstrated three haplotypes and a dN/dS ratio of 0.26 (Table 55). Analysis from fragments generated through SNP genotyping revealed synonymous mutations in Cathepsin D and paramyosin with one amino acid haplotype for all *D. gallinae* populations analysed. Across the two vitellogenin fragments amplified a total of 8 non-synonymous mutations were revealed, with two amino acid haplotypes across all populations observed, indicating high conservation of amino acid sequence, despite high nucleotide sequence diversity (Table 55).

Vaccine candidate	Fragment	Length	No. of samples	Total no. of polymorphisms	No. of synonymous polymorphisms	No. of non-synonymous polymorphisms	dN/dS Ratio	No. of amino acid haplotypes
Sanger Sequencing								
Vitellogenin	2	974bp	14	19	16	3	0.19	3
	3	629bp	11	27	21	6	0.29	10
	4	44bp	74	20	17	3	0.18	4
	6	514bp	13	10	6	4	0.67	7
	7	882bp	8	16	10	6	0.60	8
	8	806bp	15	8	7	1	0.14	2
Cathepsin D	1	836bp	6	24	19	5	0.26	3
SNP Genotyping								
Cathepsin D	A	202bp	76	1	1	0	0	1
Paramyosin	A	125bp	76	1	1	0	0	1
Vitellogenin	A	287bp	76	15	11	4	0.36	2
Vitellogenin	B	403bp	76	14	10	4	0.40	2

Table 55: *Total number of polymorphisms and non-synonymous vs synonymous mutations in Cathepsin D, Paramyosin and Vitellogenin fragments amplified from* D. gallinae *pooled samples* through Sanger sequencing and SNP genotyping

8.5 Discussion

8.5.1 Vitellogenin

A total of 91 polymorphic sites were identified across six amplified fragments of the vitellogenin coding sequence from *D. gallinae* isolates. A total of 3406bp were covered by six primer pairs out of 5538bp from the available mRNA sequence, representing 61.5% of the coding sequence. From these five were indels and the remaining 86 were SNPs. The remaining 38.5% of vitellogenin was not covered due to restricted time to optimise primer pairs for successful amplification.

8.5.1.1 Antigenic diversity of Vitellogenin

Fragments of Vitellogenin amplified through Sanger sequencing revealed between 2-10 amino acid haplotypes from the *D. gallinae* populations amplified, and two haplotypes from the 76 sequences from pooled *D. gallinae* samples analysed for SNP genotyping. Antigenic diversity was variable across the fragments, with six fragments having two to four haplotypes and two fragments having seven and ten haplotypes (fragment six and fragment three, respectively) (Table 55). The amino acid haplotype number observed indicates the existence of antigenic variation that could undermine use of Vitellogenin as a vaccine candidate. However, it should be noted that whilst variation has been observed it has not been confirmed as to whether it occurs in immunorelevant epitopes. Investigation into the relevance of antigenic haplotype to escape from an otherwise protective immune response is warranted before a conclusion can be drawn about suitability of Vitellogenin as a vaccine candidate, as if key epitopes are encoded by regions of lower diversity (e.g. fragments two, four or eight) then efficacy could be maintained across all *D. gallinae* populations.

Analysis of polymorphism in coding regions of Vitellogenin from Sanger sequencing revealed an average dN/dS ratio of 0.35, ranging from 0.14-0.67 (Table 55) and from the smaller SNP genotyping fragments a ratio of 0.36 and 0.40. The dN/ds ratio is a popular and reliable measure of evolutionary pressure on protein-coding regions, with a dN/dS<1 indicating negative selection, a dN/dS of 1 indicating neutrality and a dN/dS> indicating positive selection (765). The ratio of dN/dS is expected to exceed one when natural selection promotes changes in the coding sequence (765). Results of dN/dS ratios for the vitellogenin fragments sequenced suggest that the Vitellogenin gene is under negative selection, supporting hypothesis two, that vaccine candidate genes will be under purifying selection, encouraging use as a vaccine candidate.

8.5.1.2 Conserved regions of Vitellogenin

Vitellogenin genes share similar structural motifs, such as an N-terminal lipid binding domain (LPD_N), the unknown functional region (DUF1943), a von Willebrand factor type D similar domain (vWD), cleavage sites (R/KXXR) and C-terminal GLCG domain (766, 767). The GL/ICG motif is considered the most highly conserved Vitellogenin domain and is essential for the oligomerization of the vertebrate Vitellin (766, 768). The motif has been mutated in different species of *Acarina*, such as the Citrus red mite, *Panonychus citri*, *T. urticae* and the soft tick *Ornithodoros moubata*, which contain the GLCG motif (768, 769), the hard tick *D. variabilis*, which contains GLCS residues (767), and the parasitic mite *V. destructor*, which contains GVCG residues. A BLASTP of the *D. gallinae* amino acid reference sequence revealed three putative conserved domains: the LPD_N, vWD and the DUF1943 region as seen other species (Figure 54). Fragments two, three and four correspond to the LPD_domain, with low numbers of non-synonymous mutations and amino acid haplotypes demonstrated in fragments two and four (Table 55), but a higher number of amino acid haplotypes seen in fragment three. The Conserved Domains Database did not produce any hits for cleavage sits (K/XXR) or the C-terminal GLCG domain. Research on the Citrus red mite, *P. citri*, focusing on the cDNA of vitellogenin (*PcVg1*) and Vitellogenin receptor (*PcVgR*) demonstrated high conservation of Vg primary structures when compared to other species. They also proposed that the vitellogenin receptor, as a single gene transcript, has higher conservation than Vitellogenin, which is encoded by multiple genes (766). Work on the wolf spider *Paradosa pseudoannulata* also revealed conserved structural characteristics of Vitellogenin with respect to other insect and arachnid (770). BLAST results demonstrated over 50% identity with seven other mite species, six of which belonging to the same Order as *D. gallinae* with identity observed for Vitellogenin 1, 3 and 6. Multiple Vg genes are described from *D. variabilis* (Acari: Ixodidae) (708, 767), *H. longicornis* (Acari: Ixodidae) (697), *Neoseiulus cucumeris* (Acari: Phytoseiidae) (771), *V.destructor* (Acari: Varroidae) (772), *R. microplus* (Acari: Ixodidae) (773), *Tetranychus cinnabarinus* (Acari: Tetranychidae) (774), *Rhipicephalus appendiculatus* (Acari: Ixodidae) (775), *Tropilaelaps mercedesae* (Acari, Laelapidae) (776) and *T. urticae* (Acari: Tetranychidae) (769).

Comparison of 10 paralogous vitellogenins from ten insect species revealed a high degree of amino acid composition conservation despite changes in nucleotide sequences (777). The authors concluded that the conservation of amino acid composition is achieved through a balanced loss and gain of each amino acid residue that subjects insect vitellogenin's to an unusual system of purifying selection whereby amino acid composition is conserved over sequence composition. They hypothesised that this selection could arise from the nutritional needs of an embryo selecting for maintenance of amino acid balance (777). Results from *D.*

gallinae vitellogenin fragments indicate similarity to that seen in insects, with a greater number of synonymous mutations observed compared to non-synonymous, suggesting that amino acid conservation is favoured (Table 55). However, in some fragments high amino acid haplotypes were observed (Fragment three and four) (Table 55).

8.5.1.3 Nucleotide and haplotype diversity in Vitellogenin

Population diversity is described mainly by haplotype diversity (Hd) and nucleotide diversity (π) (778). Fragment four, corresponding to nucleotide positions 1581-2023, had the highest nucleotide diversity (Table 50), but fragment three, corresponding to positions 1093-1719 (138bp cross-over), showed the highest average number of nucleotide differences. The sliding window analysis demonstrated the nucleotide diversity of fragment four was primarily attributed to four clusters of polymorphisms, with the other positions demonstrating little or no variability (Figure 53). Fragments three, six and seven had the same number of haplotypes as isolates sequenced, attributing each isolate to an individual haplotype and giving a haplotype diversity score of 1. BLASTP identified three putative conserved domains, as discussed in 1.5.1.1, with fragments two, three and four corresponding to the N-terminal lipid binding domain (relating to nucleotide positions 93-2130). This implies that a similar system as described in insects could be in place in *D. gallinae*, whereby amino acid composition is conserved over sequence composition. Fragment four was the most intensively sampled fragment, with 51 haplotypes identified from 74 sequences (Table 50). Fragment four covers the von Willebrand factor type D (nucleotide positions 4461-4965bp), which could explain the low diversity observed throughout most of the fragment, as it is known to be highly conserved due to serving an essential function. The dN/dS ratio for fragment four was demonstrated to be 0.67, indicating purifying selection, however seven amino acid haplotypes were identified, suggesting that conservation of the von Willebrand factor type D is lower than expected in *D. gallinae*. Whilst haplotype diversity is high, low nucleotide diversity values indicated that there are only small differences between haplotypes. This combination of high haplotype diversity and low nucleotide diversity, as demonstrated here, can be indicative of rapid population expansion from a small effective population size (347). In *D. gallinae* population expansion can occur rapidly, with the ability to complete its lifecycle within 7-10 days (20), allowing fortnightly doubling of population size under optimal conditions. Other plausible explanations for rapid population expansion include introducing new flocks to empty hen houses, where a few *D. gallinae* have survived cleaning of poultry houses or poor acaricide/cleaning procedures on farms allow for survival (347). A sufficient level of gene flow amongst different *D. gallinae* populations can slow down or even prevent the process of geographic differentiation and this results in leaving a signature of little population structure over a large geographic area (779).

8.5.1.4 Neutrality tests: Vitellogenin

Under the theory of neutrality, the means of θw (nucleotide polymorphism) and π (nucleotide diversity) should equal each other, meaning that the expected Tajima's D value for populations adhering to a standard neutral model will be zero (351). As for Tajima's D, expected values for Li's F and D would also be zero (351). Results from Tajima's D and Fu and Li's F and D test revealed no significant deviation from 0 for all fragments of vitellogenin analysed, indicating neutrality in the vitellogenin gene (Tables 25-53). Across the vitellogenin fragments on a sliding window analysis, both Tajima's D and Fu and Li's D and F test followed an almost identical pattern (Figure 55).

8.5.2 Cathepsin D

Comparison of cDNA amplicons revealed 24 mutations between base pair positions 189-1025 of the Cat-D reference (Table 45). SNP genotyping revealed 42.1% of samples had the alternative allele for the marker analysed in Cat-D, with no additional variation observed in the entire fragment amplified across the UK and the rest of the Europe.

8.5.2.1 Antigenic diversity of Cathepsin D

Analysis of cDNA amplified by Sanger sequencing revealed a dN/dS ratio of 0.26 (Table 55) with three haplotypes present from six samples. Analysis of the shorter sequences generated for SNP genotyping revealed a synonymous mutation and one haplotype present across all 75 pooled *D. gallinae* populations sampled. A dN/dS ratio <1 is indicative of Cathepsin-D being under negative/purifying selection, however it should be noted that a small number of samples were analysed. Sequencing of a greater number of *D. gallinae* population across the entire coding length of Cathepsin-D would give greater clarification and provide a more robust dN/dS ratio. Overall, the results indicate that antigenic diversity in Cathepsin D is low, with high conservation of amino-acids, suggesting that vaccination with Cathepsin-D could be promising, resulting in high efficacy is all haplotypes were covered.

8.5.2.2 Nucleotide diversity, haplotype diversity and neutrality tests: Cathepsin D

As observed with Vitellogenin, high haplotype diversity and low genetic diversity was observed for Cathepsin D, supporting the notion of rapid population expansion and enough gene flow amongst *D. gallinae* populations leaving signatures of little population structure. Tajima's D and Fu and Li's D and F statistic did not significantly deviate from 0, indicating that Cathepsin D is under neutral selection.

8.5.2.3 Amplification of Cathepsin D

Amplification of Cat-D from gDNA proved complicated with failure to successfully amplify any of the primer pairs designed. Amplification of cDNA utilising the same primer pairs proved successful, implying that complications with amplifying gDNA could be related to presence of introns in Cat-D, multiple copies of Cat-D present in *D. gallinae* or poor gDNA quality. Research indicates three isoforms of Cat-D (IrCat-D1-3) are encoded within the *I. ricinus genome*, with each playing distinct and central roles in development and physiology (780, 781). Differences in these tick isoforms were observed in both primary and tertiary protein structures, as well as differential expression patterns in tick developmental stages (780). A number of factors are known to influence PCR success including: the primer length, PCR buffer reagents, simple repeats in the primer sequence, GC contents of primer and template, stable secondary structure of product and primer sequences etc (782-785). Thermodynamically stable secondary binding sites are known to leave to ambiguous sequences being produced (785). All the primers designed for amplification of Cat-D had a weak or no secondary structure when designed on Sigma's Oligocalculator (Table 43). Other parameters that influence primer quality reduce the amount of single stranded primers available in the anneal step (loop formation/self-complementarity) or through modification of the primer elongation efficiency (785). It is known that utilisation of non-optimal primers can result in amplification of undesired regions or no amplification at all (782).

8.5.3 Paramyosin and Tropomyosin

SNP genotyping results revealed the alternative allele present in 19.7% of samples analysed for paramyosin (Table 54). No additional mutations were observed in the amplified paramyosin fragment, with 100% amino acid conservation across all samples indicating that antigenic diversity in this gene is low, as a result of serving an essential function as discussed in the introduction (7.1.2.3). With proven success at inducing *D. gallinae* mortality and low genetic diversity, it could offer a strong vaccine candidate, but further amplification of the entire gene would be required to fully understand the level of diversity present. Due to COVID-19 causing restrictions to laboratory access it was not possible to amplify paramyosin from individual or pooled *D. gallinae* samples.

8.5.4 Complications with amplification via PCR

PCR failure was observed primarily for Cat-D but also for vitellogenin. Individual samples would amplify successfully for some primer pairs but fail to produce a result for others. Commonly PCR problems are dealt with through optimisation of PCR conditions, including concentrations of reagents in PCR buffer or modification of primer annealing temperature etc. (782, 786). One

study focusing on amplifying COI and ITS fragments from individual mites observed PCR failure in a small proportion of their samples (235). They noted that highest amplification success was achieved when fresh mites were used for DNA extraction as DNA extraction was less successful in dry specimens. They noted that for some samples both COI and ITS regions failed to amplify, indicating absence of PCR-quality DNA, but that in the case of ten samples, amplification was successful for ITS but failed for COI and they attribute this to variation in primer annealing sites (235). Gradient PCRs were performed for all primer pairs to identify the optimal running temperature, but these were only completed with a select few samples. It is possible that variation in primer annealing sites as suggested by Oines and Brannstrom (235) could be one explanation for PCR failure.

8.6 CONCLUSION

Analysis by Sanger sequencing and SNP genotyping of Cathepsin D, Paramyosin and Vitellogenin from *D. gallinae* samples demonstrated low antigenic diversity in all three. Results indicate that all three remain valid vaccine candidates against *D. gallinae*, but additional sequencing and investigation is required to assess the relevance of the limited antigenic diversity detected in Vitellogenin.

9 GENERAL DISCUSSION

Dermanyssus gallinae causes a significant economic loss to the European poultry industry with estimates of €230 million lost per annum (67). This cost has mostly been attributed to production losses, higher feed conversion ratios and requirements for control *(22)*. Current estimates for worldwide prevalence of *D. gallinae* range between 20-90% of layer chicken farms (21, 24, 26, 30, 265), dependant on the production system utilised and the country. As well as economic losses, *D. gallinae* represents a serious welfare concern for chickens when infestation levels range from moderate to high, including an increase in irritation, cannibalism, anaemia, restlessness, feather pecking and even hen mortality (35). *Dermanyssus gallinae* appears to demonstrate considerably plasticity regarding host specificity and has been shown to be capable of feeding on mammals, including humans (41). A role in transmission of multiple pathogenic agents with veterinary as well as zoonotic relevance has also been suggested for *D. gallinae*. Currently, isolation from *D. gallinae* mites has been demonstrated for some pathogenic agents (e.g. Newcastle disease virus) (58), whilst transmission has been confirmed for others (e.g. *Pasteurella multocida*) (787).

Current methods of control aimed at *D. gallinae* are mostly ineffective and widespread resistance to acaricides has been demonstrated across Europe (25, 115). In 2017, Exzolt was released by MSD Animal Health (266), using fluralaner as a novel control against *D. gallinae*, however it is generally considered expensive (as noted by UK farmers in the questionnaire undertaken here) and resistance is anticipated to develop in the future. Novel control strategies are urgently required in order to reduce the health, welfare and economic losses that *D. gallinae* incurs and one consideration is the development of a suitable vaccine or novel drugs. Vaccination has been shown to be a feasible approach for controlling *D. gallinae*, although optimal antigens and strategies for delivery are yet to be determined. As part of the optimisation process, and to safeguard the efficacy of any new drugs or novel control methods that are developed in the future, a more in-depth knowledge of population structure and genetic diversity will be invaluable for *D. gallinae*.

The main aim of the thesis was to gain an understanding of *D. gallinae* population structure and genetic diversity. Exploring diversity across *D. gallinae* genomes for the first time, I have focused on the relevance of diversity to current and future forms of control including variation that contributes to acaricide resistance and antigenic diversity that may influence the efficacy and longevity of future subunit/recombinant vaccines. In order to achieve this, five main objectives were set out: (1) collection and processing of *D. gallinae* from field locations; (2) identification and validation of genetic markers for *D. gallinae*; (3) genome-wide genetic analysis to assess *D. gallinae* population structure and regional variation; (4) studying the occurrence of genotypes

that have been associated with acaricide resistance to pyrethroids and (5) assessment of genetic diversity at loci encoding anti-*D. gallinae* vaccine candidates .

Sampling of *D. gallinae* was conducted across the UK and other countries in Europe. From the UK, a total of 24 farms were sampled, covering all four countries and 18 counties. Three farms were sampled on multiple occasions to assess variation over time. All farms were layer systems. Most UK farms were free-range (88%, with 36% having organic status). The remaining 12% of farms used an intensive production system. From the rest of Europe, samples were received from 16 countries, covering 82 individual farms, with the average number of farms per country five.

9.1 POPULATION STRUCTURE AND GENETIC DIVERSITY OF D. GALLINAE

Whilst one might hypothesise that a 'natural' population of *D. gallinae*, that is one that occurs naturally in the wild and is not limited to commercial poultry production, would be expected to adhere to Hardy-Weinberg equilibrium (HWE), populations of *D. gallinae* sampled exclusively from commercial poultry houses, as studied here, will likely be subject to migration and selection that violate the HWE assumptions. Previous research has proven distinct genetic differences between *D. gallinae* sampled from wild birds and *D. gallinae* sampled from commercial poultry, with minimal cross-breeding between the two populations observed (45, 235).

Genome-wide analysis of 145 SNP markers from 75 pooled *D. gallinae* samples revealed high spatial genetic diversity in *D. gallinae* populations, with no conserved haplotypes detected at more than one location or on more than one occasion. Significant linkage disequilibrium was observed across all populations (except intensive layer farms, likely due to small sample size), indicating historical and on-going admixture between *D. gallinae* populations. Amplification of a fragment of the COI gene from an overlapping sample set demonstrated evidence for genetic diversities in *D. gallinae* distributed across Europe, with phylogenetic analysis providing further support for historial or ongoing international and intranational movement of *D. gallinae*, supporting descriptions from previous research (237). Additionally, sampling individual mites from the same barn at a single time point revealed intra-farm variation in the UK and Greece, providing further support for on-going admixture at farm level. Sanger sequencing of the locus encoding the vaccine candidate, Vitellogenin, revealed a similar picture to that of COI, with high haplotype diversity, low nucleotide diversity and admixture between UK and European populations. Assessing the occurrence of mutations previously associated with pyrethroid resistance revealed distinct VGSC profiles across countries and high levels of polymorphism, with the UK demonstrating the highest incidence of mutations.

The combination of high haplotype diversity and low nucleotide diversity, as observed across multiple different research strands in this thesis, can be indicative of rapid population contraction and expansion, resulting in multiple small but overlapping populations (347). Results demonstrating variable VGSC profiles both within the UK and across the rest of Europe indicate a high level of pre-existing genetic diversity in *D. gallinae* with intra-farm variation observed in COI fragments suggesting high interbreeding. In combination, high pre-existing genetic diversity and interbreeding in conjunction with rapid population contractions (i.e. during flock turnover), followed by rapid expansion (i.e. colonisation throughout the flock cycle) will result in low nucleotide diversity, due to a slow mutation rate (since population contraction occurred), but high haplotype diversity resulting from on-going hybridisation. Genetic drift can influence the genetic structure of small populations through increased differentiation (605), meaning that small founding *D. gallinae* populations in farms may have differentiated through genetic drift during population colonisation and expansion, shifting allelic frequencies in such a way to form individual haplotypes. Whilst trade across Europe of poultry, and poultry related equipment, could cause unintentional movement of *D. gallinae* populations, and thus introduction and mixing of new alleles into farms, it might be at an insufficient level of gene flow to permit homogenisation of allele frequencies. This would result in the spatial genetic diversity observed across *D. gallinae* populations sampled (604).

Dermanyssus gallinae spends the majority of its lifecycle residing in cracks and crevices, only going onto the host to feed. In combination with the commercial laying farms analysed, it is possible to hypothesise the starting populations for these farms were small, coming from infested hens or contaminated equipment. Commercial layer farms also go through regular flock changes (~ every 72 weeks), featuring de-population and, usually, deep cleaning. Consequentially, *D. gallinae* populations experience repeated rounds of rapid population decline and subsequent population expansion. As *D. gallinae* can reside in very small crevices inside poultry houses, there is potential for a small number of individuals to survive the cleaning process. Poor acaricidal or cleaning application could increase the number of individuals that survive further (347). Rapid population expansion from these founding individuals can be supplemented through admixture with new individuals bought into the farm during flock replenishment, facilitated by the ability of *D. gallinae* to complete its lifecycle within 7-10 days (20), allowing fortnightly doubling of population size under optimal conditions. These founder events would further reduce effective population sizes and lead to higher genetic differentiation in *D. gallinae* via genetic drift.

Results from the questionnaire provided to UK farmers during sampling indicated a wide range of production systems, housing styles, organic status, type and combination of control measures

and flock sizes. The heterogeneity of control measures against *D. gallinae*, and the adherence by staff to upholding good hygiene practices, could both play a role in the genetic diversity of *D. gallinae* observed. The individuality by which each farm adopts these management practices and control measures could lead to little or no consistent directional selection on specific haplotypes across the UK as populations of *D. gallinae* at each farm are under specific local selection pressures related to the chosen control measure(s) in place.

9.1.1 Population structure and genetic diversity across production systems

Significant differences in genetic diversity within a fragment of the COI gene were seen across UK production systems, with a smaller number of more diverse genotypes observed in free-range farms compared to intensive farms. This was possibly due to multiple factors, including outdoor access in free-range farms providing a larger environment for transmission and parasite persistance (102), higher *D. gallinae* populations recorded in free-range farms (103, 104), differential use of control measures and selection pressures present at each system and greater difficulty cleaning free-range systems. These factors could permit increased admixture of *D. gallinae* within farms potentially increasing opportunities for recombination events. In contrast to a single locus (i.e. COI), results from the SNP genotyping panel revealed minimal differences in the genetic diversity of free-range and intensive production systems, with high haplotype numbers observed across both systems and phylogenetic analysis of UK production systems revealing no differentiation in phylogeny based on production system. Whilst free-range production systems offer greater transmission and persistance opportunities for *D. gallinae*, all production systems are subjected to the same migration and selection that causes deviation from HWE, through rapid population decline and expansion from flock turnaround and continual unintentional mixing of new *D. gallinae* into existing populations.

9.1.2 Temporal changes in genetic diversity in *D. gallinae* populations

Results from SNP genotyping of 145 markers revealed distinct haplotypes were observed from different visits for all farms visited on more than one occasion (UK6, UK7 and UK11), indicating that changes in genetic diversity of *D. gallinae* populations occur over time. Network analysis of these farms demonstrated different phylogenetic patterns for each farm, indicating that changes in the genetic diversity of *D. gallinae* populations are not only related to individual farms, but that they do not appear to follow a universal pattern. Temporal changes in the occurrence of mutations relating to pyrethroid resistance were also demonstrated for UK6, UK7 and UK 11, with changes in both the occurrence and the number of substitutions observed from multiple sampling events. One plausible explanation for changes in *D. gallinae* populations over time is the introduction of new alleles into the population and changes in population size during

flock turn around, resulting in a shift in allelic frequency. Individuality in temporal changes to genetic diversity in farms is likely related to differences in several management factors including use of control measures against *D. gallinae*, production system and housing system utilised, hygiene practices in place and adherence by staff, and flock size.

9.2 FUTURE WORK

To provide further elucidation of the population structure of *D. gallinae* and genetic diversity in populations across the world, inclusion of countries outside of Europe would be optimal. *Dermanyssus gallinae* has been detected in mulitple countries in Africa (28, 262) and Asia (261, 263), and inclusion of *D. gallinae* populations from these countries would enable a global view of genetic diversity and population structure rather than one limited to Europe.

Further investigation into mutations associated with pyrethroid resistance putatively identified in *D. gallinae* populations utilising individual mite DNA extracts would provide an opportunity to provide clarification about whether mutations are truly 'fixed' in a population (i.e. only alternative alleles present in all individual mites) and the role they play in resistance. Additionally, research into the mutations, including the F1357 'pyrethroid sensing residue' identified in 65.7% of UK populations and 21.6% of European farms, would provide validation of their role or lack of- in resistance to pyrethroids. Information gained surrounding these the functional association of these markers and any additional mutations could be utilised in the development of molecular diagnostics for resistance. Farmers would have the option to use this tool to help inform the optimal selection of acaricide(s) for use in controlling *D. gallinae* on their farm, reducing the cost of control through avoidance of non-optimal acaricides and increasing control efficacy.

Whilst temporal changes in genetic diversity and frequency of mutations relating to pyrethroid resistance were observed, research focusing on temporal changes in *D. gallinae* would aid in clarification of the effects of rapid population decline and expansion (e.g. bottle neck events) associated with flock turnaround. Multiple samples taken at regular time intervals (e.g. once a week or month) from all barns present at a single farm, timed to represent changes at the farm, for example, changing of flocks, could enable a deeper understanding. Extraction of individual mite DNA from many individuals, or multiple pools of *D. gallinae*, taken for every time point would also permit a more in-depth analysis.

Overcoming the limitations posed by the current *D. gallinae* genome assembly with improved assembly, curation and annotation, would be optimal for future research into *D. gallinae*. An improved genome sequence assembly would facilitate further research into population

structure and genetic diversity of *D. gallinae* populations at a deeper level, as well as providing the opportunity for research focused on specific genes or sets of genes, such as vaccine candidates. The future use of low coverage whole genome sequencing from field populations could improve genome-wide genetic analysis, assessing variation at the 32,599 SNPs identified here and expanding resources.

9.3 Contribution to the field of research

The research included in this thesis provides the first genome-wide investigation of genetic diversity in *D. gallinae* populations across the UK and the rest of Europe, as well as identification of an informative panel of SNP markers. A subset of these markers has been validated, with proven efficacy for use in genetics studies. A total panel of 32,599 SNPs were identified through the GATK pipeline and these could be used by other researchers to develop panels of SNPs relating to specific genes, when annotations permit, or to conduct further genome-wide genetic analysis. Additionally, sample collection of *D. gallinae* from 24 UK farms and 82 European farms (spanning 16 countries) also facilitated the first study mapping mutations associated with pyrethroid resistance (592) with further SNP genotyping of these markers providing a more in depth analysis. Whilst research into the mortality inducing effects of vaccine candidates has been investigated (182, 186, 188, 192), this thesis provides a preliminary investigation into antigenic diversity for four anti-*D gallinae* loci, that have not previously been studied.

The knowledge gained from this thesis regarding genetic diversity in *D. gallinae* provides an informative basis for the future development of novel control measures, including vaccines and acaricides. High spatial genetic diversity in *D. gallinae* population indicate that resistance to control methods can be developed quickly, and thus, vaccine candidates and target genes relating to novel drugs developed should be sequenced across populations to determine genetic and antigenic (amino acid) diversity. This will provide insight crucial for development of control measures, informing on likely efficacy and the likelihood of resistance development. Highly conserved genes serving essential functions are less likely to develop resistance due to selection pressure on these genes to maintain sequence conservation. In terms of existing control measures, knowledge of mutations relating to pyrethroid resistance could be utilised by farmers to reduce the economic impact of control by avoiding use of pyrethroids in populations demonstrating resistant genotypes.

REFERENCES

1. O'Neill BC, Dalton M, Fuchs R, Jiang L, Pachauri S, Zigova K. Global demographic trends and future carbon emissions. Proc Natl Acad Sci U S A. 2010;107(41):17521-6.
2. Smil V. Worldwide transformation of diets, burdens of meat production and opportunities for novel food proteins. Enzyme Microb Tech. 2002;30(3):305-11.
3. Conan A, Goutard FL, Sorn S, Vong S. Biosecurity measures for backyard poultry in developing countries: a systematic review. Bmc Vet Res. 2012;8.
4. Council BP. How the sector works 2016 [Available from: http://www.britishpoultry.org.uk/how-the-sector-works/.
5. EggInfo. Egg Production. 2019.
6. EggInfo. Egg Production 2016.
7. Commission E. Laying Hens 2019 [Available from: https://ec.europa.eu/food/animals/welfare/practice/farm/laying_hens_en.
8. Info E. Egg Production 2019 [Available from: https://www.egginfo.co.uk/egg-facts-and-figures/production.
9. Zenner L, Bon G, Dernburg A, Lubac S, Chauve C. Preliminary studies of the monitoring of *Dermanyssus gallinae* in free-range poultry farms. Br Poult Sci. 2003;44(5):781-2.
10. Ruff MD. Important parasites in poultry production systems. Veterinary Parasitology. 1999;84(3-4):337-47.
11. Axtell RC, Arends JJ. Ecology and management of arthropod pests of poultry. Annu Rev Entomol. 1990;35:101-26.
12. Hoy MA. The predatory mite Metaseiulus occidentalis: mitey small and mitey large genomes. Bioessays. 2009;31(5):581-90.
13. Capinera J. Encyclopedia of Entomology New York Springer 2008.
14. Proctor H, Owens I. Mites and birds: diversity, parasitism and coevolution. Trends Ecol Evol. 2000;15(9):358-64.
15. Brooke MD. The Effect of Allopreening on Tick Burdens of Molting Eudyptid Penguins. Auk. 1985;102(4):893-5.
16. Roy L, Chauve CM. Historical review of the genus Dermanyssus Duges, 1834 (Acari: Mesostigmata: Dermanyssidae). Parasite. 2007;14(2):87-100.

17. Knee W, Proctor H, Galloway T. Survey of nasal mites (Rhinonyssidae, Ereynetidae, and Turbinoptidae) associated with birds in Alberta and Manitoba, Canada. Can Entomol. 2008;140(3):364-79.
18. Roy L, Dowling AP, Chauve CM, Lesna I, Sabelis MW, Buronfosse T. Molecular phylogenetic assessment of host range in five Dermanyssus species. Exp Appl Acarol. 2009;48(1-2):115-42.
19. Chauve C. The poultry red mite Dermanyssus gallinae (De Geer, 1778): current situation and future prospects for control. Vet Parasitol. 1998;79(3):239-45.
20. Koziatek S, Sokół R. *Dermanyssus gallinae* still poses a serious threat for the rearing of laying hens. Polish Journal of Natural Sciences. 2015;30(4):451-63.
21. Cencek T. Prevalence of Dermanyssus gallinae in poultry farms in silesia region in Poland. Bulletin of the Veterinary Institute in Pulawy. 2003;47(2):465-9.
22. Sparagano O, Pavlicevic A, Murano T, Camarda A, Sahibi H, Kilpinen O, et al. Prevalence and key figures for the poultry red mite Dermanyssus gallinae infections in poultry farm systems. Exp Appl Acarol. 2009;48(1-2):3-10.
23. Hoglund J, Nordenfors H, Uggla A. Prevalence of the poultry red mite, Dermanyssus gallinae, in different types of production systems for egg layers in Sweden. Poult Sci. 1995;74(11):1793-8.
24. Guy JH, Khajavi M, Hlalel MM, Sparagano O. Red mite (Dermanyssus gallinae) prevalence in laying units in Northern England. Br Poult Sci. 2004;45 Suppl 1:S15-6.
25. Marangi M, Morelli V, Pati S, Camarda A, Cafiero MA, Giangaspero A. Acaricide residues in laying hens naturally infested by red mite Dermanyssus gallinae. PLoS One. 2012;7(2):e31795.
26. Fiddes MD, Le Gresley S, Parsons DG, Epe C, Coles GC, Stafford KA. Prevalence of the poultry red mite (Dermanyssus gallinae) in England. Vet Rec. 2005;157(8):233-5.
27. Oh S, Noh G, Yi S, Do Y, Kim E, Yoo J. Molecular epidemiological characterization of poultry red mite (*Dermanyssus gallinae*) collected from Korea. Korean Journal of Veterinary Service. 2019;42(3):161-7.
28. Gharbi M, Sakly N, Darghouth MA. Prevalence of Dermanyssus gallinae (Mesostigmata: Dermanyssidae) in industrial poultry farms in North-East Tunisia. Parasite. 2013;20:41.

29. Waap H, Nunes T, Mul MF, Gomes J, Bartley K. Survey on the prevalence of Dermanyssus gallinae in commercial laying farms in Portugal. Avian Pathol. 2019;48(sup1):S2-S9.

30. Hamidi A, Sherifi K, Muji S, Behluli B, Latifi F, Robaj A, et al. Dermanyssus gallinae in layer farms in Kosovo: a high risk for Salmonella prevalence. Parasit Vectors. 2011;4:136.

31. Meyer-Kuhling B, Pfister K, Muller-Lindloff J, Heine J. Field efficacy of phoxim 50% (ByeMite) against the poultry red mite Dermanyssus gallinae in battery cages stocked with laying hens. Vet Parasitol. 2007;147(3-4):289-96.

32. Immediato D, Camarda A, Iatta R, Puttilli MR, Ramos RA, Di Paola G, et al. Laboratory evaluation of a native strain of Beauveria bassiana for controlling Dermanyssus gallinae (De Geer, 1778) (Acari: Dermanyssidae). Vet Parasitol. 2015;212(3-4):478-82.

33. Sparagano OA, George DR, Harrington DW, Giangaspero A. Significance and control of the poultry red mite, Dermanyssus gallinae. Annu Rev Entomol. 2014;59:447-66.

34. Nordenfors H, Hoglund J, Uggla A. Effects of temperature and humidity on oviposition, molting, and longevity of Dermanyssus gallinae (Acari: Dermanyssidae). J Med Entomol. 1999;36(1):68-72.

35. Kilpinen O, Roepstorff A, Permin A, Norgaard-Nielsen G, Lawson LG, Simonsen HB. Influence of Dermanyssus gallinae and Ascaridia galli infections on behaviour and health of laying hens (Gallus gallus domesticus). Br Poult Sci. 2005;46(1):26-34.

36. Maurer V. The dynamics of *Dermanyssus gallinae* (Acari: Dermanyssidae) populations interacting with laying hens and the predatory mite *Cheyletus eruditus* (Acari: Cheyletidae). ETH Zürich: Swiss Federal Institute of Technology Zurich; 1993.

37. Wojcik AR, Grygon-Franckiewicz B, Zbikowska E, Wasielewski L. [Invasion of Dermanyssus gallinae (De Geer, 1778) in poultry farms in the Torun region]. Wiad Parazytol. 2000;46(4):511-5.

38. Arkle S, Guy JH, Sparagano O. Immunological effects and productivity variation of red mite (Dermanyssus gallinae) on laying hens-implications for egg production and quality. World Poultry Sci J. 2006;62(2):249-57.

39. Kowalski A, Sokol R. Influence of Dermanyssus gallinae (poultry red mite) invasion on the plasma levels of corticosterone, catecholamines and proteins in layer hens. Pol J Vet Sci. 2009;12(2):231-5.

40. Sokol R, Szkamelski A, Barski D. Influence of light and darkness on the behaviour of Dermanyssus gallinae on layer farms. Pol J Vet Sci. 2008;11(1):71-3.

41. Valiente Moro C, De Luna CJ, Tod A, Guy JH, Sparagano OA, Zenner L. The poultry red mite (Dermanyssus gallinae): a potential vector of pathogenic agents. Exp Appl Acarol. 2009;48(1-2):93-104.

42. Roy L, Chauve C, Buronfosse T. Contrasted ecological repartition of the Northern Fowl Mite *Ornithonyssus sylviarum* (Mesostigmata : Macronyssidae) and the Chicken Red Mite *Dermanyssus gallinae* (Mesostigmata : Dermanyssidae). Acarologia. 2010;50(2):207-19.

43. Roy L, Dowling A, Chauve CM, Buronfosse T. Delimiting species boundaries within D*ermanyssus* Dugès, 1834 (Acari: Mesostigmata) using a total evidence approach. Molecular Phylogenetics and Evolution. 2009;50:446-70.

44. Roy L, Buronfosse T. Using mitochondrial and nuclear sequence data for disentangling population structure in complex pest species: a case study with Dermanyssus gallinae. PLoS One. 2011;6(7):e22305.

45. Brannstrom S, Morrison DA, Mattsson JG, Chirico J. Genetic differences in internal transcribed spacer 1 between Dermanyssus gallinae from wild birds and domestic chickens. Med Vet Entomol. 2008;22(2):152-5.

46. Lima-Barbero JF, Diaz-Sanchez S, Sparagano O, Finn RD, de la Fuente J, Villar M. Metaproteomics characterization of the alphaproteobacteria microbiome in different developmental and feeding stages of the poultry red mite Dermanyssus gallinae (De Geer, 1778). Avian Pathol. 2019;48(sup1):S52-S9.

47. Moro CV, Desloire S, Chauve C, Zenner L. Detection of Salmonella sp. in Dermanyssus gallinae using an FTA filter-based polymerase chain reaction. Med Vet Entomol. 2007;21(2):148-52.

48. Huong CT, Murano T, Uno Y, Usui T, Yamaguchi T. Molecular detection of avian pathogens in poultry red mite (Dermanyssus gallinae) collected in chicken farms. J Vet Med Sci. 2014;76(12):1583-7.

49. Shirnov FB, Ibragiomova AI, Misirov ZG. The dissemination of the fowl-pox by the mite *Dermanyssus gallinae*. Veterinarya. 1972;4:48-9.

50. Smith MG, Blattner RJ, Heys FM. The Isolation of the St. Louis Encephalitis Virus from Chicken Mites (Dermanyssus Gallinae) in Nature. Science. 1944;100(2599):362-3.

51. Smith MG, Blattner RJ, Heys FM. St. Louis encephalitis; infection of chicken mites, dermanyssus gallinae, by feeding on chickens with viremia; transovarian passage of virus into the second generation. J Exp Med. 1946;84:1-6.

52. Chamberlain RW, Sikes RK, Sudia WD. Attempted Laboratory Infection of Bird Mites with the Virus of St-Louis Encephalitis. American Journal of Tropical Medicine and Hygiene. 1957;6(6):1047-53.

53. Wegner Z. Laboratory study on some parasitic hematophagous arthropods as possible subsidiary links of the biocenosis of tick-borne encephalitis. Bull Inst Marit Trop Med Gdynia. 1976;27:75-85.

54. Durden LA, Linthicum KJ, Monath TP. Laboratory Transmission of Eastern Equine Encephalomyelitis Virus to Chickens by Chicken Mites (Acari, Dermanyssidae). Journal of Medical Entomology. 1993;30(1):281-5.

55. Chamberlain RW, Sikes RK. Laboratory Investigations on the Role of Bird Mites in the Transmission of Eastern and Western Equine Encephalitis. American Journal of Tropical Medicine and Hygiene. 1955;4(1):106-18.

56. Durden LA, Linthicum KJ, Turell MJ. Mechanical transmission of Venezuelan equine encephalomyelitis virus by hematophagous mites (Acari). J Med Entomol. 1992;29(1):118-21.

57. Hoffmann G. Vogelmilben als Lastlinge, Krankheitserzeuger und Vektoren bei Mensch und Nutztier. Dtsch Tier¨arztl Wschr. 1987;95:7-10.

58. Arzey G. Mechanism of spread of Newcastle disease. New South Wales Agriculture and Fisheries Bulletin. 1990;41:12.

59. Zeman P, Stika V, Skalka B, Bartik M, Dusbabek F, Lavickova M. Potential role of Dermanyssus gallinae De Geer, 1778 in the circulation of the agent of pullurosis-typhus in hens. Folia Parasitol (Praha). 1982;29(4):371-4.

60. Petrov D. Study of Dermanyssus gallinae as a carrier of *Pasteurella multocida*. Vet Med Nauki. 1975;12(5):32-6.

61. Chirico J, Eriksson H, Fossum O, Jansson D. The poultry red mite, Dermanyssus gallinae, a potential vector of Erysipelothrix rhusiopathiae causing erysipelas in hens. Med Vet Entomol. 2003;17(2):232-4.

62. Zemskaya AA, Pchelkina AA. Gamasoid mites and Q fever. 1967. In: Problemy Parazitologii [Internet]. Kiev; [258-9].

63. Grebenyuk RV, Chirov PA, Kadysheva AM. The role of wild animals and blood-sucking arthropods in the epizootiology of infection with Listerioza 1972.

64. Valiente Moro C, Thioulouse J, Chauve C, Normand P, Zenner L. Bacterial taxa associated with the hematophagous mite Dermanyssus gallinae detected by 16S rRNA PCR amplification and TTGE fingerprinting. Res Microbiol. 2009;160(1):63-70.

65. George DR, Finn RD, Graham KM, Mul MF, Maurer V, Moro CV, et al. Should the poultry red mite Dermanyssus gallinae be of wider concern for veterinary and medical science? Parasit Vectors. 2015;8:178.

66. Van Emous R. Wage war agaisnt the red mite! . Poultry Interational. 2005;44:26-33.

67. Van Emous R. Verwachtte schade bloedluis 21 miljoen euro 2017 [Available from: https://www.pluimveeweb.nl/artikel/163578-verwachtte-schade-bloedluis-21-miljoen-euro/.

68. Mul MF, Koenraadt CJ. Preventing introduction and spread of Dermanyssus gallinae in poultry facilities using the HACCP method. Exp Appl Acarol. 2009;48(1-2):167-81.

69. Cosoroaba I. Observation d'invasions massives par *Dermanyssus gallinae* (De Geer 1778) chez les poules élevées en batterie en Roumanie. Revue de Médecine Vétérinaire. 2001;152(1):89-96.

70. Sleeckx N, Van Gorp S, Koopman R, Kempen I, Van Hoye K, De Baere K, et al. Production losses in laying hens during infestation with the poultry red mite Dermanyssus gallinae. Avian Pathol. 2019;48(sup1):S17-S21.

71. Roy L, Chauve C, Buronfosse T. Contrasted ecological repartition of the Northern Fowl Mite *Ornithonyssus sylviarum* (Mesostigmata : Macronyssidae) and the Chicken Red Mite *Dermanyssus gallinae* (Mesostigmata : Dermanyssidae). Acarologia 2010;50(2):207-19.

72. Cafiero MA, Camarda A, Circella E, Santagada G, Schino G, Lomuto M. Pseudoscabies caused by Dermanyssus gallinae in Italian city dwellers: a new setting for an old dermatitis. J Eur Acad Dermatol Venereol. 2008;22(11):1382-3.

73. Williams RW. An infestation of a human habitation by Dermanyssus gallinae (Degeer, 1778) (Acarina: Dermanyssidae) in New York City resulting in sanguisugent attacks upon the occupants. Am J Trop Med Hyg. 1958;7(6):627-9.
74. Sulzberger MB, Kaminstein I. Avian itch mites as a cause of human dermatoses - Canary birds mites responsible for two groups of cases in New York. Arch Dermatol Syph. 1936;33(1):60-72.
75. Shaw JW, Pommerening RA. Avian Mite Dermatitis (Gamasoidosis) - Report of a Case. Arch Dermatol Syph. 1950;61(3):466-9.
76. Rockwell E. Dermatitis due to Dermanyssus gallinae of pigeons. AMA Arch Derm Syphilol. 1953;68(1):82.
77. Deoreo GA. Pigeons acting as vector in acariasis caused by Dermanyssus gallinae (DeGeer, 1778). AMA Arch Derm. 1958;77(4):422-9.
78. Winkler A. Endemie durch Vogelmilben (*Dermanyssus gallinae*) in einem Krankenhaus. Dermatol Wochenschr. 1967;16:458-9.
79. Rosen S, Yeruham I, Braverman Y. Dermatitis in humans associated with the mites Pyemotes tritici, Dermanyssus gallinae, Ornithonyssus bacoti and Androlaelaps casalis in Israel. Med Vet Entomol. 2002;16(4):442-4.
80. Akdemir C, Gulcan E, Tanritanir P. Case report: Dermanyssus gallinae in a patient with pruritus and skin lesions. Turkiye Parazitol Derg. 2009;33(3):242-4.
81. Haag-Wackernagel D, Bircher AJ. Ectoparasites from feral pigeons affecting humans. Dermatology. 2010;220(1):82-92.
82. Prins M, Go IH, van Dooren-Greebe RJ. [Parasitic pruritus: bird mite zoonosis]. Ned Tijdschr Geneeskd. 1996;140(51):2550-2.
83. Baselga E, Drolet BA, Esterly NB. Avian mite dermatitis. Pediatrics. 1996;97(5):743-5.
84. Fuentes M, Sainz-Elipes S, Saez-Duran S, Galan-Puchades M. Human ectoparasitism due to the poultry red mite, Dermanyssus gallinae, in the city of Valencia (Spain) and its surroundings. Parasitology. 2009;68:188-91.
85. Collgros H, Iglesias-Sancho M, Aldunce MJ, Exposito-Serrano V, Fischer C, Lamas N, et al. Dermanyssus gallinae (chicken mite): an underdiagnosed environmental infestation. Clin Exp Dermatol. 2013;38(4):374-7.

86. Abdigoudarzi M, Mirafzali MS, Belgheiszadeh H. Human Infestation with Dermanyssus gallinae (Acari: Dermanyssidae) in a Family Referred with Pruritus and Skin Lesions. J Arthropod Borne Dis. 2014;8(1):119-23.
87. Pampiglione S, Pampiglione G, Pagani M, Rivasi F. [Persistent scalp infestation by Dermanyssus gallinae in an Emilian country-woman]. Parassitologia. 2001;43(3):113-5.
88. MacInnis AJ. How parasites find hosts: some thoughts on the inception of host-parasite integration. In: Kennedy CR, editor. Ecological Aspects of Parasitology. The Netherlands: North-Holland Publishing Co.; 1976. p. 3-20.
89. Sutcliffe JF. Distance Orientation of Biting Flies to Their Hosts. Insect Sci Appl. 1987;8(4-6):611-6.
90. Kilpinen O. Activation of the poultry red mite, Dermanyssus gallinae (Acari: Dermanyssidae), by increasing temperatures. Exp Appl Acarol. 2001;25(10-11):859-67.
91. Kilpinen O, Mullens BA. Effect of food deprivation on response of the mite, Dermanyssus gallinae, to heat. Med Vet Entomol. 2004;18(4):368-71.
92. Koenraadt CJ, Dicke M. The role of volatiles in aggregation and host-seeking of the haematophagous poultry red mite Dermanyssus gallinae (Acari: Dermanyssidae). Exp Appl Acarol. 2010;50(3):191-9.
93. Entrekin DL, Oliver JH, Jr. Aggregation of the chicken mite, Dermanyssus gallinae (acari: Dermanyssidae). J Med Entomol. 1982;19(6):671-8.
94. Zeman P. Surface skin lipids of birds--a proper host kairomone and feeding inducer in the poultry red mite, Dermanyssus gallinae. Exp Appl Acarol. 1988;5(1-2):163-73.
95. Maurer V, Bieri M, Folsch DW. Host-Finding of Dermanyssus-Gallinae in Poultry-Houses. Arch Geflugelkd. 1988;52(5):209-15.
96. Kirkwood A. Some observations on the feeding habits of the poultry mites *Dermanyssus gallinae* and *Liponyssus sylviarum* Entomologia Experimentalis et Applicata 1968;11:315-20.
97. Kirkwood AC. In vitro feeding of Dermanyssus gallinae. Exp Parasitol. 1971;29(1):1-6.
98. Flochlay AS, Thomas E, Sparagano O. Poultry red mite (Dermanyssus gallinae) infestation: a broad impact parasitological disease that still remains a significant challenge for the egg-laying industry in Europe. Parasites & Vectors. 2017;10.

99. Affairs DfEFaR. United Kingdom Egg Statistics – Quarter 3, 2020. In: Statistics N, editor. 2020. p. 4.

100. Kirkwood AC. A trap perch for the control of the poultry red mite (Dermanyssus gallinae). Br Poult Sci. 1965;6(1):73-8.

101. Nordenfors H, Chirico J. Evaluation of a sampling trap for Dermanyssus gallinae (Acari: Dermanyssidae). J Econ Entomol. 2001;94(6):1617-21.

102. Pavlicevic A, Pavlovic I, Stajkovic N. Method for early detection of poultry red mite D*ermanyssus gallinae* (De Geer, 1778). Biotechnology in Animal Husbandry. 2007;23(3-4):119-27.

103. Zenner L, Bon G, Chauve C, Nemoz C, Lubac S. Monitoring of Dermanyssus gallinae in free-range poultry farms. Exp Appl Acarol. 2009;48(1-2):157-66.

104. Cox M, De Baere K, Vervaet E, Zoons J, Fiks-Van T. Red mites: monitoring method and treatment. Italy: Cervia; 2009. p. 83.

105. Mul MF, van Riel JW, Meerburg BG, Dicke M, George DR, Groot Koerkamp PW. Validation of an automated mite counter for Dermanyssus gallinae in experimental laying hen cages. Exp Appl Acarol. 2015;66(4):589-603.

106. Mul M. Control methods for *Dermanyssus gallinae* in systems for laying hens: results of an international seminar. Worlds Poultry Science Journal. 2009;65:589-99.

107. Abbas RZ, Colwell DD, Iqbal Z, Khan A. Acaricidal drug resistance in poultry red mite (Dermanyssus gallinae) and approaches to its management. World Poultry Science Association 2014;70:113-24.

108. Beugnet F, Chauve C, Gauthey M, Beert L. Resistance of the red poultry mite to pyrethroids in France. Vet Rec. 1997;140(22):577-9.

109. Chirico J, Tauson R. Traps containing acaricides for the control of Dermanyssus gallinae. Vet Parasitol. 2002;110(1-2):109-16.

110. Moser R, Raffaelli R, Thilmany-McFadden D. Consumer Preferences for Fruit and Vegetables with Credence-Based Attributes: A Review. Int Food Agribus Man. 2011;14(2):121-41.

111. Hamscher G, Priess B, Nau H. Determination of phoxim residues in eggs by using high-performance liquid chromatography diode array detection after treatment of stocked housing facilities for the poultry red mite (Dermanyssus gallinae). Anal Chim Acta. 2007;586(1-2):330-5.

112. News B. Fipronil egg scandal: What we know 2017 [updated 11/08/2017. Available from: https://www.bbc.co.uk/news/world-europe-40878381.

113. World P. Fipronil scandal: Defra launches testing regime on British egg 2017 [updated 04/08/2017 Available from: https://www.poultryworld.net/UK/Articles/2017/8/Fipronil-scandal-Defra-launches-testing-regime-on-British-egg-167598E/.

114. Zeman P. Encounter the Poultry Red Mite Resistance to Acaricides in Czechoslovak Poultry-Farming. Folia Parasitologica. 1987;34(4):369-73.

115. Marangi M, Cafiero MA, Capelli G, Camarda A, Sparagano OA, Giangaspero A. Evaluation of the poultry red mite, Dermanyssus gallinae (Acari: Dermanyssidae) susceptibility to some acaricides in field populations from Italy. Exp Appl Acarol. 2009;48(1-2):11-8.

116. Nordenfors H, Hoglund J, Tauson R, Chirico J. Effect of permethrin impregnated plastic strips on Dermanyssus gallinae in loose-housing systems for laying hens. Vet Parasitol. 2001;102(1-2):121-31.

117. Georghiou GP, Taylor CE. Genetic and biological influences in the evolution of insecticide resistance. J Econ Entomol. 1977;70(3):319-23.

118. Feyereisen R. Molecular biology of insecticide resistance. Toxicol Lett. 1995;82-83:83-90.

119. Taylor M, Feyereisen R. Molecular biology and evolution of resistance of toxicants. Mol Biol Evol. 1996;13(6):719-34.

120. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Mol Biol. 2005;14(1):3-8.

121. Feyereisen R. Insect cytochrome P450. In: Latrou K, Gilbert LI, Gill SS, editors. Comprehensive Molecular Insect Science. 5. Oxford: Elsevier; 2005. p. 1-77.

122. Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD, Russel RJ. Biochemical genetics and genomics of insect esterases. In: Latrou K, Gilbert LI, Gill SS, editors. Comprehensive Molecular Insect Science. 5. Oxford: Elsevier; 2005. p. 309-82.

123. Li X, Schuler MA, Berenbaum MR. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu Rev Entomol. 2007;52:231-53.

124. Maurer V, Perler E, Heckendorn F. In vitro efficacies of oils, silicas and plant preparations against the poultry red mite Dermanyssus gallinae. Exp Appl Acarol. 2009;48(1-2):31-41.

125. Steenberg T, Kilpinen O. Synergistic interaction between the fungus Beauveria bassiana and desiccant dusts applied against poultry red mites (Dermanyssus gallinae). Exp Appl Acarol. 2014;62(4):511-24.

126. Ebeling W. Sorptive dusts for pest control. Annu Rev Entomol. 1971;16:123-58.

127. Kirkwood AC. Sulphonamides for the experimental control of the poultry red mite Dermanyssus gallinae (De Geer, 1778). Ann Trop Med Parasitol. 1974;68(1):125-8.

128. Kilpinen O, Steenberg T. Inert dusts and their effects on the poultry red mite (Dermanyssus gallinae). Exp Appl Acarol. 2009;48(1-2):51-62.

129. Subramanyam B, Roesli, R. Inert dusts. In: Subramanyam B, Hagstrum, D.W., editor. Alternatives to

pesticides in stored-product IPM. Doordrecht: Kluwer Academic Publishers; 2000. p. 321-80.

130. Maurer V, Perler, E. Silicas for control of the poultry red mites *Dermanyssus gallinae*. Joint Organic Congress; Odense, Denmark2006.

131. Arthur FH. Toxicity of diatomaceous earth to red flour beetles and confused flour beetles (Coleoptera: Tenebrionidae): effects of temperature and relative humidity. J Econ Entomol. 2000;93(2):526-32.

132. Schulz J, Berk J, Suhl J, Schrader L, Kaufhold S, Mewis I, et al. Characterization, mode of action, and efficacy of twelve silica-based acaricides against poultry red mite (Dermanyssus gallinae) in vitro. Parasitol Res. 2014;113(9):3167-75.

133. Lubac S, Dernburg A, Bon G, Chauve C, Zenner L. Problématique et pratiques d'élevages en poulespondeuses dans le sud est de la France contre les nuisibles: poux rouges et mouches. In: ITAVI I, AFSSA, editor. 5emes journées de la recherche avicole, Tours, France. Tours2003. p. 101-4.

134. Zoons J. The effect of light programs on red mite (*Dermanyssus gallinae*) in battery cage housing. In: Perry GC, editor. Welfare of the laying hen CABI Poultry Science Symposium. 27. Wallingford: CABI; 2004. p. 416.

135. Charmillot PJ, Hofer D, Pasquier D. Attract and kill: a new method for control of the codling moth Cydia pomonella. Entomologia Experimentalis Et Applicata. 2000;94(2):211-6.

136. Stetter J, Lieb F. Innovation in crop protection: Trends in research. Angew Chem Int Edit. 2000;39(10):1724-44.

137. Vargas RI, Miller NW, Stark JD. Field trials of spinosad as a replacement for naled, DDVP, and malathion in methyl eugenol and cue-lure bucket traps to attract and kill male oriental fruit flies and melon flies (Diptera : Tephritidae) in Hawaii. Journal of Economic Entomology. 2003;96(6):1780-5.

138. Tinzaara W, Gold CS, Dicke M, Van Huis A, Nankinga CM, Kagezi GH, et al. The use of aggregation pheromone to enhance dissemination of Beauveria bassiana for the control of the banana weevil in Uganda. Biocontrol Sci Techn. 2007;17(1-2):111-24.

139. Kim SI, Yi JH, Tak JH, Ahn YJ. Acaricidal activity of plant essential oils against Dermanyssus gallinae (Acari : Dermanyssidae). Veterinary Parasitology. 2004;120(4):297-304.

140. Tavassoli M, Ownag A, Pourseyed SH, Mardani K. Laboratory evaluation of three strains of the entomopathogenic fungus Metarhizium anisopliae for controlling Dermanyssus gallinae. Avian Pathol. 2008;37(3):259-63.

141. Lundh J, Wiktelius D, Chirico J. Azadirachtin-impregnated traps for the control of Dermanyssus gallinae. Vet Parasitol. 2005;130(3-4):337-42.

142. Pritchard J, Kuster T, George D, Sparagano O, Tomley F. Impeding movement of the poultry red mite, Dermanyssus gallinae. Vet Parasitol. 2016;225:104-7.

143. Quintero MT, Acevedo AC. Studies on deep litter mites in poultry farms in Mexico. In: Griffiths DA, Bowman CE, editors. Acarology VI. 1. Chichester: Ellis Horwood Ltd; 1984. p. 629-34.

144. Brady J. The mites of poultry litter: observations on the bionomics of common species, with a species list for England and Wales. Journal of Applied Ecology 1970;7:331-48.

145. Brady J. Litter mites and their effects on poultry. World Poultry Sci J. 1970;26:658-68.

146. Abo-Taka SM, editor Mites Inhabiting poultry farms in Egypt Acarology IX; 1996; Columbus Ohio Bioloigcal Survery

147. Lesna I, Wolfs P, Faraji F, Roy L, Komdeur J, Sabelis MW. Candidate predators for biological control of the poultry red mite Dermanyssus gallinae. Exp Appl Acarol. 2009;48(1-2):63-80.

148. Lesna I, Sabelis MW, van Niekerk TG, Komdeur J. Laboratory tests for controlling poultry red mites (Dermanyssus gallinae) with predatory mites in small 'laying hen' cages. Exp Appl Acarol. 2012;58(4):371-83.

149. Ali W, George DR, Shiel RS, Sparagano OA, Guy JH. Laboratory screening of potential predators of the poultry red mite (Dermanyssus gallinae) and assessment of Hypoaspis miles performance under varying biotic and abiotic conditions. Vet Parasitol. 2012;187(1-2):341-4.

150. Thompson GD, Dutton R, Sparks TC. Spinosad - a case study: an example from a natural products discovery programme. Pest Management Science. 2000;56(8):696-702.

151. Anastas P, Kirchchoff M, Williamson T. Green Chemistry awards: spinosad - a new natural product for insect control. Green Chemistry 1999;1:88.

152. Rhimi W, Ben Salem I, Camarda A, Saidi M, Boulila A, Otranto D, et al. Chemical characterization and acaricidal activity of Drimia maritima (L) bulbs and Dittrichia viscosa leaves against Dermanyssus gallinae. Vet Parasitol. 2019;268:61-6.

153. Borges LMF, de Sousa LAD, Barbosa CD. Perspectives for the use of plant extracts to control the cattle tick Rhipicephalus (Boophilus) microplus. Rev Bras Parasitol V. 2011;20(2):89-96.

154. Adenubi OT, McGaw LJ, Eloff JN, Naidoo V. In vitro bioassays used in evaluating plant extracts for tick repellent and acaricidal properties: A critical review. Veterinary Parasitology. 2018;254:160-71.

155. Kim SI, Yi JH, Tak JH, Ahn YJ. Acaricidal activity of plant essential oils against Dermanyssus gallinae (Acari: Dermanyssidae). Vet Parasitol. 2004;120(4):297-304.

156. George DR, Smith TJ, Sparagano OA, Guy JH. The influence of 'time since last blood meal' on the toxicity of essential oils to the poultry red mite (Dermanyssus gallinae). Vet Parasitol. 2008;155(3-4):333-5.

157. George DR, Smith TJ, Shiel RS, Sparagano OA, Guy JH. Mode of action and variability in efficacy of plant essential oils showing toxicity against the poultry red mite, Dermanyssus gallinae. Vet Parasitol. 2009;161(3-4):276-82.

158. Radsetoulalova I, Hubert J, Hampel D, Lichovníková M. Active components of essential oils as acaricides against *Dermanyssus gallinae*. British Poultry Science. 2019.

159. Kim SI, Na YE, Yi JH, Kim BS, Ahn YJ. Contact and fumigant toxicity of oriental medicinal plant extracts against Dermanyssus gallinae (Acari: Dermanyssidae). Vet Parasitol. 2007;145(3-4):377-82.

160. Ghrabi-Gammar Z, George DR, Daoud-Bouattour A, Jilani IB, Ben Saad-Limam S, Sparagano OAE. Screening of essential oils from wild-growing plants in Tunisia for their

yield and toxicity to the poultry red mite, Dermanyssus gallinae. Ind Crop Prod. 2009;30(3):441-3.

161. Na YE, Kim SI, Bang HS, Kim BS, Ahn YJ. Fumigant toxicity of cassia and cinnamon oils and cinnamaldehyde and structurally related compounds to Dermanyssus gallinae (Acari: Dermanyssidae). Vet Parasitol. 2011;178(3-4):324-9.

162. Gorji SF, Gorji SF, Rajabloo M. The field efficacy of garlic extract against Dermanyssus gallinae in layer farms of Babol, Iran. Parasitology Research. 2014;113(3):1209-13.

163. Nechita IS, Poirel MT, Cozma V, Zenner L. The repellent and persistent toxic effects of essential oils against the poultry red mite, Dermanyssus gallinae. Vet Parasitol. 2015;214(3-4):348-52.

164. Tabari MA, Youssefi MR, Barimani A, Araghi A. Carvacrol as a potent natural acaricide against Dermanyssus gallinae. Parasitol Res. 2015;114(10):3801-6.

165. Barimani A, Youssefi MR, Tabari MA. Traps containing carvacrol, a biological approach for the control of Dermanyssus gallinae. Parasitol Res. 2016;115(9):3493-8.

166. Steenberg T, Kilpinen, O. Fungus infection of the chicken mite *Dermanyssus gallinae*. IOBC/WPRS Bull. 2003;26:23-5.

167. de Faria MR, Wraight SP. Mycoinsecticides and Mycoacaricides: A comprehensive list with worldwide coverage and international classification of formulation types. Biol Control. 2007;43(3):237-56.

168. Tavassoli M, Allymehr, M., Pourseyed, S.H., Ownag, A., Bernousi, I., Mardani, K., Ghorbanzadegan, M., Shokrpoor, S. . Field bioassay of *Metharhizium anisopliae* strains to control the poultry red mite *Dermanyssus gallinae*. Veterinary Parasitology. 2011;178:374-8.

169. Lord JC. Desiccant dusts synergize the effect of Beauveria bassiana (Hyphomycetes: Moniliales) on stored-grain beetles. J Econ Entomol. 2001;94(2):367-72.

170. Lord JC. Low humidity, moderate temperature, and desiccant dust favor efficacy of Beauveria bassiana (Hyphomycetes : Moniliales) for the lesser grain borer, Rhyzopertha dominica (Coleoptera : Bruchidae). Biol Control. 2005;34(2):180-6.

171. Akbar W, Lord JC, Nechols JR, Howard RW. Diatomaceous earth increases the efficacy of Beauveria bassiana against Tribolium castaneum larvae and increases Conidia attachment. Journal of Economic Entomology. 2004;97(2):273-80.

172. Thompson SR, Brandenburg RL. Effect of combining imidacloprid and diatomaceous earth with Beauveria bassiana on mole cricket (Orthoptera : Gryllotalpidae) mortality. Journal of Economic Entomology. 2006;99(6):1948-54.

173. Luz C, Rodrigues J, Rocha LFN. Diatomaceous earth and oil enhance effectiveness of Metarhizium anisopliae against Triatoma infestans. Acta Trop. 2012;122(1):29-35.

174. Hunt DWA, Borden JH, Rahe JE, Whitney HS. Nutrient-Mediated Germination of Beauveria-Bassiana Conidia on the Integument of the Bark Beetle Dendroctonus-Ponderosae (Coleoptera, Scolytidae). J Invertebr Pathol. 1984;44(3):304-14.

175. Crampton A, Vanniasinkam T. Parasite vaccines: the new generation. Infect Genet Evol. 2007;7(5):664-73.

176. Cox FE. Designer vaccines for parasitic diseases. Int J Parasitol. 1997;27(10):1147-57.

177. Good MF, Stanisic D, Xu H, Elliott S, Wykes M. The immunological challenge to developing a vaccine to the blood stages of malaria parasites. Immunol Rev. 2004;201:254-67.

178. de la Fuente J, Almazan C, Canales M, Perez de la Lastra JM, Kocan KM, Willadsen P. A ten-year review of commercial vaccine performance for control of tick infestations on cattle. Anim Health Res Rev. 2007;8(1):23-8.

179. de la Fuente J, Kopacek P, Lew-Tabor A, Maritz-Olivier C. Strategies for new and improved vaccines against ticks and tick-borne diseases. Parasite Immunol. 2016;38(12):754-69.

180. de la Fuente J, Contreras M. Tick vaccines: current status and future directions. Expert Rev Vaccines. 2015;14(10):1367-76.

181. Contreras M, Alberdi, P., Fernández de Mera, I.G., Krull, C., Nijhof, A., Villar, M.,and de la Fuente. J. . Vaccinomics approach to the identification of candidate protective antigens for the control of tick vector infestations and *Anaplasma phagocytophilum* infection. Frontiers in Cellular and Infection Microbiology 2017;7:360.

182. Bartley K, Turnbull F, Wright HW, Huntley JF, Palarea-Albaladejo J, Nath M, et al. Field evaluation of poultry red mite (Dermanyssus gallinae) native and recombinant prototype vaccines. Veterinary Parasitology. 2017;244:25-34.

183. Contreras M, Villar M, Artigas-Jeronimo S, Kornieieva L, Mytrofanov S, de la Fuente J. A reverse vaccinology approach to the identification and characterization of Ctenocephalides felis candidate protective antigens for the control of cat flea infestations. Parasite Vector. 2018;11.

184. Artigas-Jeronimo S, Villar M, Cabezas-Cruz A, Valdes JJ, Estrada-Pena A, Alberdi P, et al. Functional Evolution of Subolesin/Akirin. Front Physiol. 2018;9:1612.

185. de la Fuente J, Estrada-Pena A. Why New Vaccines for the Control of Ectoparasite

Vectors Have Not Been Registered

and Commercialized? Vaccines. 2019(7):75.

186. Bartley K, Wright HW, Huntley JF, Manson ED, Inglis NF, McLean K, et al. Identification and evaluation of vaccine candidate antigens from the poultry red mite (Dermanyssus gallinae). Int J Parasitol. 2015;45(13):819-30.

187. Bartley K, Nisbet AJ, Offer JE, Sparks NH, Wright HW, Huntley JF. Histamine release factor from Dermanyssus gallinae (De Geer): characterization and in vitro assessment as a protective antigen. Int J Parasitol. 2009;39(4):447-56.

188. Bartley K, Huntley JF, Wright HW, Nath M, Nisbet AJ. Assessment of cathepsin D and L-like proteinases of poultry red mite, Dermanyssus gallinae (De Geer), as potential vaccine antigens. Parasitology. 2012;139(6):755-65.

189. Lima-Barbero JF, Contreras M, Bartley K, Price DRG, Nunn F, Sanchez-Sanchez M, et al. Reduction in Oviposition of Poultry Red Mite (Dermanyssus gallinae) in Hens Vaccinated with Recombinant Akirin. Vaccines (Basel). 2019;7(3).

190. Lima-Barbero JF, Contreras M, Mateos-Hernandez L, Mata-Lorenzo FM, Triguero-Ocana R, Sparagano O, et al. A vaccinology Approach to the Identification and Characterization of Dermanyssus Gallinae Candidate Protective Antigens for the Control of Poultry Red Mite Infestations. Vaccines (Basel). 2019;7(4).

191. Harrington D, Robinson K, Guy J, Sparagano O. Characterization of the immunological response to Dermanyssus gallinae infestation in domestic fowl. Transbound Emerg Dis. 2010;57(1-2):107-10.

192. Wright HW, Bartley K, Huntley JF, Nisbet AJ. Characterisation of tropomyosin and paramyosin as vaccine candidate molecules for the poultry red mite, Dermanyssus gallinae. Parasit Vectors. 2016;9(1):544.

193. Nordenfors H, Hoglund J. Long term dynamics of dermanyssus gallinae in relation to mite control measures in aviary systems for layers. Br Poult Sci. 2000;41(5):533-40.

194. Willadsen P, Kemp DH. Vaccination with 'concealed' antigens for tick control. Parasitol Today. 1988;4(7):196-8.

195. Harrington DW, Robinson K, Sparagano OA. Immune responses of the domestic fowl to Dermanyssus gallinae under laboratory conditions. Parasitol Res. 2010;106(6):1425-34.

196. Harrington D, Din HM, Guy J, Robinson K, Sparagano O. Characterization of the immune response of domestic fowl following immunization with proteins extracted from Dermanyssus gallinae. Vet Parasitol. 2009;160(3-4):285-94.

197. Wright HW, Bartley K, Nisbet AJ, McDevitt RM, Sparks NH, Brocklehurst S, et al. The testing of antibodies raised against poultry red mite antigens in an in vitro feeding assay; preliminary screen for vaccine candidates. Exp Appl Acarol. 2009;48(1-2):81-91.

198. Willadsen P. Anti-tick vaccines. Parasitology. 2004;129 Suppl:S367-87.

199. Arkle S, Harrington D, Kaiser P, Rothwell L, De Luna C, George D, et al. Immunological control of the poultry red mite. Ann N Y Acad Sci. 2008;1149:36-40.

200. Nisbet AJ, Huntley JF, Mackellar A, Sparks N, McDevitt R. A house dust mite allergen homologue from poultry red mite Dermanyssus gallinae (De Geer). Parasite Immunol. 2006;28(8):401-5.

201. Price DRG, Kuster T, Oines O, Oliver EM, Bartley K, Nunn F, et al. Evaluation of vaccine delivery systems for inducing long-lived antibody responses to Dermanyssus gallinae antigen in laying hens. Avian Pathol. 2019;48(sup1):S60-S74.

202. Harrington D, Canales M, de la Fuente J, de Luna C, Robinson K, Guy J, et al. Immunisation with recombinant proteins subolesin and Bm86 for the control of Dermanyssus gallinae in poultry. Vaccine. 2009;27(30):4056-63.

203. McDevitt R, Nisbet AJ, Huntley JF. Ability of a proteinase inhibitor mixture to kill poultry red mite, Dermanyssus gallinae in an in vitro feeding system. Vet Parasitol. 2006;141(3-4):380-5.

204. Lima-Barbero JF, Villar M, Hofle U, De la Fuente J. Challenges for the Control of Poultry Red Mite (*Dermanyssus gallinae*) 2020 [

205. Wikel SK. Modulation of the host immune system by ectoparasitic arthropods - Blood-feeding and tissue-dwelling arthropods manipulate host defenses to their advantage. Bioscience. 1999;49(4):311-20.

206. Titus RG, Bishop JV, Mejia JS. The immunomodulatory factors of arthropod saliva and the potential for these factors to serve as vaccine targets to prevent pathogen transmission. Parasite Immunology. 2006;28(4):131-41.

207. Geary TG, Thompson DP. Development of antiparasitic drugs in the 21st century. Veterinary parasitology. 2003;115(2):167-84.

208. Willadsen P. Vaccination against ectoparasites. Parasitology. 2006;133:S9-S25.

209. Schultz J, Maisel S, Gerlach D, Muller T, Wolf M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA. 2005;11(4):361-4.

210. Cruickshank RH. Molecular markers for the phylogenetics of mites and ticks. Systematic & Applied Acarology. 2002;7:3-14.

211. Navajas M, Fenton B. The application of molecular markers in the study of diversity in acarology: A review. Experimental and Applied Acarology. 2000;24(10-11):751-74.

212. Galtier N, Nabholz B, Glemin S, Hurst GDD. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol. 2009;18(22):4541-50.

213. Fenton B, Malloch G, Moxey E. Analysis of eriophyid mite rDNA internal transcribed spacer sequences reveals variable simple sequence repeats. Insect Mol Biol. 1997;6(1):23-32.

214. Delrieu-Trottin E, Maynard J, Planes S. Endemic and widespread coral reef fishes have similar mitochondrial genetic diversity. P Roy Soc B-Biol Sci. 2014;281(1797).

215. Fujisawa T, Vogler AP, Barraclough TG. Ecology has contrasting effects on genetic variation within species versus rates of molecular evolution across species in water beetles. P Roy Soc B-Biol Sci. 2015;282(1799).

216. Hebert PD, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci. 2003;270 Suppl 1:S96-9.

217. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. P Roy Soc B-Biol Sci. 2003;270(1512):313-21.

218. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014;30(12):1660-6.

219. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015;10(10):1556-66.

220. Andrews S. Fast QC: A quality control tool for high throughput sequencing 2010 [Available from: http://Available from: www.bioinformatics.babraham.ac.uk/projects/fastqc.

221. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform. 2010;11(5):473-83.

222. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

223. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

224. Nagarajan N, Pop M. Sequence assembly demystified. Nat Rev Genet. 2013;14(3):157-67.

225. Li H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. Bioinformatics. 2012;28(14):1838-44.

226. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19(6):1117-23.

227. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics. 2011;12(6):443-51.

228. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

229. Zhao M, Wang QG, Wang Q, Jia PL, Zhao ZM. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. Bmc Bioinformatics. 2013;14.

230. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21(6):974-84.

231. Zhu MF, Need AC, Han YJ, Ge DL, Maia JM, Zhu QQ, et al. Using ERDS to Infer Copy-Number Variants in High-Coverage Genomes. Am J Hum Genet. 2012;91(3):408-21.

232. Burgess STG, Bartley K, Nunn F, Wright HW, Hughes M, Gemmell M, et al. Draft Genome Assembly of the Poultry Red Mite, Dermanyssus gallinae. Microbiol Resour Announc. 2018;7(18).

233. Marangi M, Cantacessi C, Sparagano OA, Camarda A, Giangaspero A. Molecular characterization and phylogenetic inferences of Dermanyssus gallinae isolates in Italy within an European framework. Med Vet Entomol. 2014;28(4):447-52.

234. Marangi M, de Luna CJ, Cafiero MA, Camarda A, le Bouquin S, Huonnic D, et al. Phylogenetic relationship between Dermanyssus gallinae populations in European countries based on mitochondrial COI gene sequences. Exp Appl Acarol. 2009;48(1-2):143-55.

235. Oines O, Brannstrom S. Molecular investigations of cytochrome c oxidase subunit I (COI) and the internal transcribed spacer (ITS) in the poultry red mite, Dermanyssus gallinae, in northern Europe and implications for its transmission between laying poultry farms. Med Vet Entomol. 2011;25(4):402-12.

236. Roy L, Dowling AP, Chauve CM, Buronfosse T. Diversity of phylogenetic information according to the locus and the taxonomic level: an example from a parasitic mesostigmatid mite genus. Int J Mol Sci. 2010;11(4):1704-34.

237. Chu TT, Murano T, Uno Y, Usui T, Yamaguchi T. Molecular epidemiological characterization of poultry red mite, Dermanyssus gallinae, in Japan. J Vet Med Sci. 2015;77(11):1397-403.

238. Potenza L, Cafiero MA, Camarda A, La Salandra G, Cucchiarini L, Dacha M. Characterization of Dermanyssus gallinae (Acarina: Dermanissydae) by sequence analysis of the ribosomal internal transcribed spacer regions. Vet Res Commun. 2009;33(7):611-8.

239. Oh S, Noh G, Yi S, Do Y, Kim E, Yoo J. Molecular epidemiological characterization of poultry red mite (*Dermanyssus gallinae*) collected from Korea. Korean Journal of Veterinary Service 2019;42(3):161-7.

240. Normark BB. The evolution of alternative genetic systems in insects. Annual Review of Entomology. 2003;48:397-423.

241. Blackmon H, Hardy NB, Ross L. The evolutionary dynamics of haplodiploidy: Genome architecture and haploid viability. Evolution. 2015;69(11):2971-8.

242. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman TL, et al. Sex Determination: Why So Many Ways of Doing It? Plos Biol. 2014;12(7).

243. Normark BB. The Evolution of Insect Mating Systems2014.

244. Norton RA, Kethley JB, Johnston D, E. , O'Conner BM. Phylogenetic perspectives on genetic systems and reproductive modes of mites. In: Wrensch DL, Ebbert MA, editors. Evolution and Diversity of Sex Ratio in Haplodiploid Insects and Mites. New York: Chapman & Hall; 1993. p. 8-99.

245. de la Filia AG, Bain SA, Ross L. Haplodiploidy and the reproductive ecology of Arthropods. Curr Opin Insect Sci. 2015;9:36-43.

246. Zahavi A. Mate selection-a selection for a handicap. J Theor Biol. 1975;53(1):205-14.

247. Boulton RA, Collins LA, Shuker DM. Beyond sex allocation: the role of mating systems in sexual selection in parasitoid wasps. Biol Rev Camb Philos Soc. 2015;90(2):599-627.

248. Reeve HK, Pfennig DW. Genetic biases for showy males: Are some genetic systems especially conducive to sexual selection? P Natl Acad Sci USA. 2003;100(3):1089-94.

249. Kirkpatrick M, Hall DW. Sexual selection and sex linkage. Evolution. 2004;58(4):683-91.

250. Smith NGC. The evolution of haplodiploidy under inbreeding. Heredity. 2000;84(2):186-92.

251. Price G. Selection and Covariance. Nature. 1970;227:520-1.

252. Sabelis MW, Nagelkerke CJ. Evolution of pseudo-arrhenotoky. Experimental and Applied Acarology 1988;4:301-18.

253. Bull JJ. An advantage for the evolution of malehaploidy and systems with similar genetic transmission. Heredity. 1979;43:361-81.

254. Bull JJ. The Evolution of Sex-determining Mechanisms. Menlo Park, Canada: Benjamin Cummings; 1983.

255. Sabelis MW, Nagelkerke CJ. Sex allocation and pseudoarrhenotoky in phytoseiid mites. In: Wrensch DL, Ebbert MA, editors. Evolution and Diversity of Sex Ratio in Haplodiploid Insects and Mites. New York: Chapman & Hall; 1993. p. 512-41.

256. Cruickshank RH, Thomas RH. Evolution of haplodiploidy in dermanyssine mites (Acari : Mesostigmata). Evolution. 1999;53(6):1796-803.

257. Oliver JH, Jr. Karyotypes and sex determination in some dermanyssid mites (Acarina: Mesostigmata). Annals of the Entomological Society of America. 1965;58(4):567-73.

258. Sigma-Aldrich. OligoEvaluator™ 2019 [Available from: http://www.oligoevaluator.com.

259. Singh V, Kumar A. PCR Primer Design. Molecular Biology Today. 2001;2(2):27-32.

260. Moore D. Purification and concentration of DNA from aqueous solutions. Curr Protoc Immunol. 2001;Chapter 10:Unit 10 1.

261. Oh SI, Do YJ, Kim E, Yi SW, Yoo JG. Prevalence of poultry red mite (Dermanyssus gallinae) in Korean layer farms and the presence of avian pathogens in the mite. Exp Appl Acarol. 2020;81(2):223-38.

262. Douifi M, Ouchene N, Hakem A, Rahal K. *Dermanyssus gallinae* in Laying Hen Houses in Algeria, Infestation Prevalence and Molecular Detection of Salmonella. Tropicultura. 2019;37(4).

263. Chu TTH, Murano T, Uno Y, Usui T, Yamaguchi T. Molecular epidemiological characterization of poultry red mite, Dermanyssus gallinae, in Japan. Journal of Veterinary Medical Science. 2015;77(11):1397-403.

264. Tomley FM, Sparagano O. Spotlight on avian pathology: red mite, a serious emergent problem in layer hens. Avian Pathol. 2018;47(6):533-5.

265. Magdas C, Chirila F, Fit N, Criste A, Baciu H. Epidemiologic study of *Dermanyssus gallinae* (*Acari:Dermanyssidae*) infestation in birds, from three localities in Cluj area. Buletin USAMV-CV. 2006:309-14.

266. MSD. Exzolt: MSD; 2020 [Available from: https://www.exzolt.com/.

267. Kirkwood AC. Anaemia in poultry infested with the red mite Dermanyssus gallinae. Vet Rec. 1967;80(17):514-6.

268. Gunnarsson E. Poultry red mites in Swedish laying hen flocks - occurrence and efficacy. Uppsala: Swedish University of Agricultural Science; 2017.

269. Grover A, Sharma PC. Development and use of molecular markers: past and present. Crit Rev Biotechnol. 2016;36(2):290-302.

270. Schlötterer C. The evolution of molecular markers - just a matter of fashion? Nature Review Genetics. 2004;5:63-9.

271. Wan QH, Wu H, Fujihara T, Fang SG. Which genetic marker for which conservation genetics issue? Electrophoresis. 2004;25(14):2165-76.

272. Khan F. Molecular Markers: An excellent tool for genetic analysis. Journal of Molecular Biomarkers & Diagnosis. 2015;6(3):233-4.

273. Sax K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics. 1932 8.

274. Wexelsen H. Linkage between quantitative and qualitative characters in barley. Hereditas. 1933;17:323-41.

275. Markert CL, Moller F. Multiple forms of enzymes: Tissue,

ontogenetic and species specific patterns. Proceedings of the National Academy of Science USA. 1959;45:753-63.

276. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995;23(21):4407-14.

277. Jaccoud D, Peng K, Feinstein D, Kilian A. Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 2001;29(4):E25.

278. Zietkiewicz E, Rafalski A, Labuda D. Genome fingerprinting by

simple sequence repeats (SSR)-anchored PCR amplification. Genomics. 1994;20:176-83.

279. Jordan SA, Humphries P. Single nucleotide polymorphism in exon 2 of the BCP gene on 7q31-q35. Hum Mol Genet. 1994;3(10):1915.

280. Akopyanz N, Bukanov NO, Westblom TU, Berg DE. PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen Helicobacter pylori. Nucleic Acids Res. 1992;20(23):6221-5.

281. Akkaya M, Bhagwat AA, Cregan PB. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics. 1992;132:1131-9.

282. Caetano-Anollés G, Bassam BJ, Gresshoff PM. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Biotechnology (N Y). 1991;9(6):553-7.

283. Adams MD, Kelly JM, Gocayne JD, Dubrick M, Polymeropoulos MH, Xioa H, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 1991;252:1651-6.
284. Williams JGK, Kublelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphism's amplified by arbitrary primers are useful as

genetic markers. Nucleic Acids Research. 1990;18:6531-5.
285. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science. 1987;235(4796):1616-22.
286. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 1990;18(24):7213-8.
287. Botstein D, White RL, Skolnick M, Davies RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. 1980;32:314-31.
288. Semagn K, Bjornstad A, Ndjiondjop MN. An overview of molecular marker methods in plants. African Journal of Biotechnology. 2006;5(25):2540-68.
289. Abdel-Mawgood AL. DNA Based Techniques for Studying Genetic Diversity. Genetic Diversity in Microorganisms. 2012:95-122.
290. Kalender R, Grob T, Regina MT, Suoniemi A, Schulman A. IRAP and REMAP: two new

retrotransposon-based DNA fingerprinting techniques. Theoretical and Applied Genetics. 1999;98:704-11.
291. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet. 1989;44(3):397-401.
292. Weber JL, May PE. Abundant class of human DNA

polymorphisms which can be typed using the

polymerase chain reaction. American Journal of Human Genetics. 1989;44:388-96.
293. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 1989;17(16):6463-71.

294. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, et al. Mining SNPs from EST databases. Genome Res. 1999;9(2):167-74.
295. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;407(6803):513-6.
296. Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K. The discovery of single-nucleotide polymorphisms--and inferences about human demographic history. Am J Hum Genet. 2001;69(6):1332-47.
297. Rota-Stabelli O, Yang Z, Telford MJ. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. Mol Phylogenet Evol. 2009;52(1):268-72.
298. Hwang UW, Kim W. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. Korean J Parasitol. 1999;37(4):215-28.
299. Shao R, Barker SC. Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. Parasitology. 2007;134(Pt 2):153-67.
300. Avise JC. Molecular Markers, Natural History, and Evolution. 2nd ed. Sunderland, MA Sinauer Associates; 2004.
301. Le TH, Blair D, McManus DP. Mitochondrial genomes of parasitic flatworms. Trends Parasitol. 2002;18(5):206-13.
302. Hu M, Gasser RB. Mitochondrial genomes of parasitic nematodes--progress and perspectives. Trends Parasitol. 2006;22(2):78-84.
303. Wolstenholme DR. Animal Mitochondrial-DNA - Structure and Evolution. International Review of Cytology-a Survey of Cell Biology. 1992;141:173-216.
304. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene-Sequences and a Compilation of Conserved Polymerase Chain-Reaction Primers. Annals of the Entomological Society of America. 1994;87(6):651-701.
305. Avise JC. Molecular Markers, Natural History and Evolution. New York: Chapman and Hall; 1994.
306. Ladoukakis ED, Zouros E. Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. J Biol Res (Thessalon). 2017;24:2.

307. Sissler M, González-Serrano LE, Westhof E. Recent Advances in Mitochondrial Aminoacyl-tRNA Synthetases and Disease. Trends Mol Med. 2017;23(8):693-708.
308. Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci U S A. 2015;112(33):10177-84.
309. Blanchard JL, Lynch M. Organellar genes: why do they end up in the nucleus? Trends Genet. 2000;16(7):315-20.
310. Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. Diversity and reductive evolution of mitochondria among microbial eukaryotes. Philos Trans R Soc Lond B Biol Sci. 2010;365(1541):713-27.
311. Makiuchi T, Nozaki T. Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. Biochimie. 2014;100:3-17.
312. John U, Lu Y, Wohlrab S, Groth M, Janouškovec J, Kohli GS, et al. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. Sci Adv. 2019;5(4):eaav1110.
313. Lavrov DV, Pett W. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. Genome Biology and Evolution. 2016;8:2896-913.
314. Dotson EM, Beard CB. Sequence and organization of the mitochondrial genome of the Chagas disease vector, Triatoma dimidiata. Insect Mol Biol. 2001;10(3):205-15.
315. Covacin C, Shao R, Cameron S, Barker SC. Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). Insect Mol Biol. 2006;15(1):63-8.
316. Shao R, Campbell NJ, Barker SC. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, Heterodoxus macropus (Phthiraptera). Mol Biol Evol. 2001;18(5):858-65.
317. Gissi C, Iannelli F, Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity. 2008;101(4):301-20.
318. Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. Mol Ecol. 2004;13(4):729-44.
319. Birky CW. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. Annu Rev Genet. 2001;35:125-48.

320. Nabholz B, Mauffrey JF, Bazin E, Galtier N, Glemin S. Determination of mitochondrial genetic diversity in mammals. Genetics. 2008;178(1):351-61.

321. Vakifahmetoglu-Norberg H, Ouchida AT, Norberg E. The role of mitochondria in metabolism and cell death. Biochem Biophys Res Commun. 2017;482(3):426-31.

322. Brown WM, Prager EM, Wang A, Wilson AC. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol. 1982;18(4):225-39.

323. Moriyama EN, Powell JR. Synonymous substitution rates in Drosophila: mitochondrial versus nuclear genes. J Mol Evol. 1997;45(4):378-91.

324. Ballard JWO, Rand DM. The population biology of mitochondrial DNA and its phylogenetic implications. Annu Rev Ecol Evol S. 2005;36:621-42.

325. Balloux F, Handley LJ, Jombart T, Liu H, Manica A. Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. Proc Biol Sci. 2009;276(1672):3447-55.

326. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science. 2004;303(5655):223-6.

327. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics. 2008;9:119.

328. Yu L, Wang X, Ting N, Zhang Y. Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. Mitochondrion. 2011;11(3):497-503.

329. Caterino MS, Cho S, Sperling FA. The current state of insect molecular systematics: a thriving Tower of Babel. Annu Rev Entomol. 2000;45:1-54.

330. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol S. 2006;37:545-79.

331. Meyer A, Zardoya R. Recent advances in the (molecular) phylogeny of vertebrates. Annu Rev Ecol Evol S. 2003;34:311-38.

332. Navajas M, Fournier D, Lagnel J, Gutierrez J, Boursot P. Mitochondrial COI sequences in mites: Evidence for variations in base composition. Insect Mol Biol. 1996;5(4):281-5.

333. Galtier N, Gouy M. Inferring phylogenies from DNA sequences of unequal base composition. Proceedings of the National Academy of Science, USA. 1995;92:11317-21.
334. Struder-Kypke MC, Lynn DH. Comparative analysis of the mitochondrial cytochrome c oxidase subunit I (COI) gene in ciliates (Alveolata, Ciliophora) and evaluation of its suitability as a biodiversity marker. Syst Biodivers. 2010;8(1):131-48.
335. Simmons RB, Weller SJ. Utility and evolution of cytochrome b in insects. Mol Phylogenet Evol. 2001;20(2):196-210.
336. Lynch M, Jarrell PE. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. Genetics. 1993;135(4):1197-208.
337. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. P Natl Acad Sci USA. 2004;101(41):14812-7.
338. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. Plos Biol. 2004;2(10):1657-63.
339. Remigio EA, Hebert PD. Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships Mol Phylogenet Evol. 2003;29:641-7.
340. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005;360:1847-57.
341. Rach J, Desalle R, Sarkar IN, Schierwater B, Hadrys H. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proc Biol Sci. 2008;275(1632):237-47.
342. Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol Ecol Notes. 2007;7(2):184-90.
343. Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN. Comprehensive DNA barcode coverage of North American birds. Mol Ecol Notes. 2007;7(4):535-43.
344. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Molecular Biology and Evolution. 2017;34:3299-302.

345. Lynch M, Crease TJ. The analysis of population survey data on DNA sequence variation. Mol Biol Evol. 1990;7(4):377-94.

346. Tajima F. The effect of change in population size on DNA polymorphism. Genetics. 1989;123(3):597-601.

347. Bhowmick B, Zhao J, Oines O, Bi T, Liao C, Zhang L, et al. Molecular characterization and genetic diversity of Ornithonyssus sylviarum in chickens (Gallus gallus) from Hainan Island, China. Parasit Vectors. 2019;12(1):553.

348. Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997;147(2):915-25.

349. Tajima F. Unbiased Estimation of Evolutionary Distance between Nucleotide Sequences. Molecular Biology and Evolution. 1993;10(3):677-88.

350. Tajima F. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics. 1989;123(3):585-95.

351. Biswas S, Akey JM. Genomic insights into positive selection. Trends Genet. 2006;22(8):437-46.

352. Kreitman M. Methods to detect selection in populations with applications to the human. Annu Rev Genomics Hum Genet. 2000;1:539-59.

353. Fu YX, Li WH. Statistical test of neutrality of mutations. Genetics. 1993;133:693-709.

354. Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. Genetics. 2008;179(1):555-67.

355. Tamura K, Kumar S. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol. 2002;19(10):1727-36.

356. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35(6):1547-9.

357. Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, McGuire G. TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. Bioinformatics. 2004;20(11):1806-7.

358. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-W9.

359. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37-48.

360. Ros VI, Breeuwer JA. Spider mite (Acari: Tetranychidae) mitochondrial COI phylogeny reviewed: host plant relationships, phylogeography, reproductive parasites and barcoding. Exp Appl Acarol. 2007;42(4):239-62.

361. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Mol Biol Evol. 1992;9:678-87.

362. Thamsborg SM, Roepstorff A, Larsen M. Integrated and biological control of parasites in organic and conventional production systems. Vet Parasitol. 1999;84(3-4):169-86.

363. Fry AJ. MILDLY DELETERIOUS MUTATIONS IN AVIAN MITOCHONDRIAL DNA: EVIDENCE FROM NEUTRALITY TESTS. Evolution. 1999;53(5):1617-20.

364. Simonsen KL, Churchill GA, Aquadro CF. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics. 1995;141(1):413-29.

365. Rand DM. Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. Conservation Biology. 1996;10:665-71.

366. Ellegren H, Galtier N. Determinants of genetic diversity. Nat Rev Genet. 2016;17(7):422-33.

367. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515(7526):261-3.

368. Hughes AR, Inouye BD, Johnson MT, Underwood N, Vellend M. Ecological consequences of genetic diversity. Ecol Lett. 2008;11(6):609-23.

369. Fisher RA. The Genetical Theory of Natural Selection. Oxford: Oxford Univerity Press; 1930.

370. Hague MT, Routman EJ. Does population size affect genetic diversity? A test with sympatric lizard species. Heredity (Edinb). 2016;116(1):92-8.

371. McNeely JA, Miller KR, Reid WV, Mittermeier RA, Werner TB. Conserving the world's biological diversity. Switzerland: International Union for Conservation of Nature and Natural Resources; 1990. 193 p.

372. Ford EB. Ecological Genetics London: Chapham & Hall; 1964.

373. Endler JA. Natural Selection in the Wild. Princeton: Princeton University Press; 1986.

374. Robert F, Pelletier J. Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. Frontier in Genetics. 2018;9:507.
375. Mccarroll SA, Hadnott TN, Perry GH, Sabeti P, C., Zody MC, Barrett JC, et al. Common deletion polymorphisms in the human genome. Nature Genetics. 2006;38:86-92.
376. Orr N, Chanock S. Common genetic variation and human disease. Adv Genet. 2008;62:1-32.
377. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. 2002;34(3):275-305.
378. Rieseberg LH. Chromosomal rearrangements and speciation. Trends Ecol Evol. 2001;16(7):351-8.
379. Bazin E, Glemin S, Galtier N. Population Size Does Not Influence

Mitochondrial Genetic Diversity

in Animals. Science. 2006;5773:572570-.
380. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. Nature Review Genetics. 2011;12:756-66.
381. Lynch M. Evolution of the mutation rate. Trends Genet. 2010;26(8):345-52.
382. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nature Review Genetics. 2009;10:195-205.
383. Kimura M. The Neutral Theory of Molecular Evolution. New York, NY, USA: Cambridge University Press; 1983.
384. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. Nat Rev Genet. 2016;17(2):93-108.
385. Cuykendall TN, Rubin MA, Khurana E. Non-coding genetic variation in cancer. Curr Opin Syst Biol. 2017;1:9-15.
386. Vidigal JA, Ventura A. The biological functions of miRNAs: lessons from in vivo studies. Trends Cell Biol. 2015;25(3):137-47.
387. Barrett LG, Thrall PH, Burdon JJ, Linde CC. Life history determines genetic structure and evolutionary potential of host-parasite interactions. Trends Ecol Evol. 2008;23(12):678-85.

388. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016;17(1):47-62.
389. Vučićević D, Schrewe H, Orom UA. Molecular mechanisms of long ncRNAs in neurological disorders. Front Genet. 2014;5:48.
390. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. Cell. 2012;148(6):1172-87.
391. Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics. 1964;49:725-38.
392. King JL, Jukes TH. Non-Darwinian evolution. Science. 1969;164(3881):788-98.
393. Kreitman M. The neutral theory is dead. Long live the neutral theory. Bioessays. 1996;18(8):678-83; discussion 83.
394. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. Annu Rev Genomics Hum Genet. 2003;4:213-35.
395. Corbett-Detig RB, Hartl DL, Sackton TB. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol. 2015;13(4):e1002112.
396. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11(10):697-709.
397. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356-69.
398. Pool JE, Hellmann I, Jensen JD, Nielsen R. Population genetic inference from genomic sequence variation. Genome Res. 2010;20(3):291-300.
399. Austerlitz F, Heyer E. Neutral Theory: From Complex Population History to Natural Selection and Sociocultural Phenomena in Human Populations. Mol Biol Evol. 2018;35(6):1304-7.
400. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A. 2014;111(17):6131-8.
401. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005;15(11):1566-75.
402. Messer PW, Petrov DA. Frequent adaptation and the McDonald-Kreitman test. Proc Natl Acad Sci U S A. 2013;110(21):8615-20.

403. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the Drosophila genome? PLoS Genet. 2009;5(6):e1000495.

404. Lewontin RC. The Genetic Basis of Evolutionary Change. New York, USA: Columbia University Press; 1974.

405. Nachman MW. Patterns of DNA variability at X-linked loci in Mus domesticus. Genetics. 1997;147(3):1303-16.

406. Amos W, Harwood J. Factors affecting levels of genetic diversity in natural populations. Philos Trans R Soc Lond B Biol Sci. 1998;353(1366):177-86.

407. Gillespie JH. Is the population size of a species relevant to its evolution? Evolution. 2001;55(11):2161-9.

408. Bazin E, Glemin S, Galtier N. Population size does not influence mitochondrial genetic diversity in animals. Science. 2006;312:570-2.

409. Andolfatto P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the Drosophila melanogaster genome. Genome Res. 2007;17(12):1755-62.

410. Macpherson JM, Sella G, Davis JC, Petrov DA. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila. Genetics. 2007;177(4):2083-99.

411. Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in Drosophila simulans. PLoS Genet. 2011;7(2):e1001302.

412. Gillespie JH. Genetic drift in an infinite population. The pseudohitchhiking model. Genetics. 2000;155(2):909-19.

413. Neher R. Genetic Draft, Selective Interference, and Population Genetics of Rapid Adaptation. Annual Review of Ecology, Evolution and Systematics. 2013;44:195-215.

414. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, et al. Natural selection on protein-coding genes in the human genome. Nature. 2005;437(7062):1153-7.

415. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8(8):610-8.

416. Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vinckenbosch N, et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genet. 2011;7(10):e1002326.

417. Chun S, Fay JC. Evidence for hitchhiking of deleterious mutations within the human genome. PLoS Genet. 2011;7(8):e1002240.

418. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134(4):1289-303.

419. Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012;190(1):5-22.

420. Hahn MW. Toward a selection theory of molecular evolution. Evolution. 2008;62(2):255-65.

421. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106(23):9362-7.

422. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods. 2014;11(3):294-6.

423. Frankham R, Ballou, J.D. and Briscoe, D.A. Introduction to Conservation Genetics. Cambridge: Cambridge University Press; 2002.

424. Weir BS. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sunderland, MA: Sinauer Associates 1996.

425. Kim S, Misra A. SNP genotyping: technologies and biomedical applications. Annu Rev Biomed Eng. 2007;9:289-320.

426. Mooney S. Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Briefings in Bioinformatics. 2005;6(1):44-56.

427. Morin P, Luikart G, Wayne R, group Sw. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution. 2004;19(4):208-16.

428. Brookes AJ. The essence of SNPs. Gene. 1999;234(2):177-86.

429. Ma ZS, Li L, Zhang YP. Defining Individual-Level Genetic Diversity and Similarity Profiles. Sci Rep. 2020;10(1):5805.

430. Duran C, Appleby N, Edwards D, Batley J. Molecular Genetic Markers: Discovery, Applications, Data Storage and

Visualisation. Current Bioinformatics. 2009;4:16-27.

431. Hagmann M. A good SNP may be hard to find. Science. 1999;285(5424):21-2.

432. Seal A, Gupta A, Mahalaxmi M, Aykkal R, Singh TR, Arunachalam V. Tools, resources and databases for SNPs and indels in sequences: a review. Int J Bioinform Res Appl. 2014;10(3):264-96.

433. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 1998;8(12):1229-31.

434. Matukumalli LK, Grefenstette JJ, Hyten DL, Choi IY, Cregan PB, Van Tassell CP. SNP-PHAGE--High throughput SNP discovery pipeline. BMC Bioinformatics. 2006;7:468.

435. Nicholson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Mühlemann O. Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci. 2010;67(5):677-700.

436. Mendell JT, Dietz HC. When the message goes awry: disease-producing mutations that influence mRNA content and performance. Cell. 2001;107(4):411-4.

437. Storer C, Pascal C, Roberts S, Templin W, Seeb L, Seeb J. Rank and Order: Evaluating the performance of SNPs for individual assignment in a non-model organism. PLoS ONE. 2012;7(11):e49018.

438. Gao J, Agrawal GK, Thelen JJ, Xu D. P3DB: a plant protein phosphorylation database. Nucleic Acids Res. 2009;37(Database issue):D960-2.

439. Lin BZ, Sasazaki S, Mannen H. Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers. Anim Sci J. 2010;81(3):281-9.

440. Selvam R, Murali N, Thiruvenkadan AK, Saravanakumar R, Ponnudurai G, Jawahar TP. Single-nucleotide polymorphism-based genetic diversity analysis of the Kilakarsal and Vembur sheep breeds. Vet World. 2017;10(5):549-55.

441. Davey JW, Davey JL, Blaxter ML, Blaxter MW. RADSeq: next-generation population genetics. Brief Funct Genomics. 2010;9(5-6):416-23.

442. Gut IG. Automation in genotyping of single nucleotide polymorphisms. Hum Mutat. 2001;17(6):475-92.

443. Kwok PY. Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet. 2001;2:235-58.

444. Rafalski JA. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science. 2002;162:329-33.

445. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. SNP discovery via 454 transcriptome sequencing. Plant J. 2007;51(5):910-8.

446. Guo M, Rupe MA, Danilevskaya ON, X. Y, Hu Z. Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. The Plant Journal. 2003;36:30-44.

447. Bray NJ, Buckland PR, Owen MJ, O'Donovan MC. Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet. 2003;113(2):149-53.

448. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. Detection of regulatory variation in mouse genes. Nat Genet. 2002;32(3):432-7.

449. Stupar RM, Springer NM. Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics. 2006;173(4):2199-210.

450. Barker G, Batley J, O' Sullivan H, Edwards KJ, Edwards D. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. Bioinformatics. 2003;19(3):421-2.

451. Buetow KH, Edmonson MN, Cassidy AB. Reliable identification of large numbers of candidate SNPs from public EST data. Nat Genet. 1999;21(3):323-5.

452. Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, et al. A general approach to single-nucleotide polymorphism discovery. Nat Genet. 1999;23(4):452-6.

453. Dantec LL, Chagné D, Pot D, Cantin O, Garnier-Géré P, Bedon F, et al. Automated SNP detection in expressed sequence tags: statistical considerations and application to maritime pine sequences. Plant Mol Biol. 2004;54(3):461-70.

454. Unneberg P, Strömberg M, Sterky F. SNP discovery using advanced algorithms and neural networks. Bioinformatics. 2005;21(10):2528-30.

455. Panitz F, Stengaard H, Hornshøj H, Gorodkin J, Hedegaard J, Cirera S, et al. SNP mining porcine ESTs with MAVIANT, a novel tool for SNP evaluation and annotation. Bioinformatics. 2007;23(13):i387-91.

456. van Oeveren J, Janssen A. Mining SNPs from DNA Sequence Data; Computational Approaches to SNP Discovery and Analysis. In: Komar A, editor. Single Nucleotide Polymorphisms Methods in Molecular Biology™ (Methods and Protocols). 578. Totowa, New Jersery: Humana Press; 2009. p. 73-91.

457. Wiltshire T, Pletcher MT, Batalov S, Barnes W, Tarantino L, Cooke MP, et al. Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. Proceedings of the National Academy of Science USA. 2003;100:3380-5.
458. Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res. 2004;14(9):1812-9.
459. Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, et al. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell. 2005;17(11):2859-72.
460. Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL. Arabidopsis map-based cloning in the post-genome era. Plant Physiol. 2002;129(2):440-50.
461. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
462. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
463. Kumar S, Banks TW, Cloutier S. SNP Discovery through Next-Generation Sequencing and Its Applications. Int J Plant Genomics. 2012;2012:831460.
464. You FM, Huo N, Deal KR, Gu YQ, Luo MC, McGuire PE, et al. Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. BMC Genomics. 2011;12:59.
465. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, et al. Tablet--next generation sequence assembly visualization. Bioinformatics. 2010;26(3):401-2.
466. Shah N, Teplitsky MV, Minovitsky S, Pennacchio LA, Hugenholtz P, Hamann B, et al. SNP-VISTA: an interactive SNP visualization tool. BMC Bioinformatics. 2005;6:292.
467. Fiume M, Williams V, Brook A, Brudno M. Savant: genome browser for high-throughput sequencing data. Bioinformatics. 2010;26(16):1938-44.
468. Schicht S, Qi WH, Poveda L, Strube C. Whole transcriptome analysis of the poultry red mite Dermanyssus gallinae (De Geer, 1778). Parasitology. 2014;141(3):336-46.

469. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018;46(W1):W537-W44.

470. Batut B, Hiltemann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, et al. Community-Driven Data Analysis Training for Biology. Cell Syst. 2018;6(6):752-8.e1.

471. team G. Germline short variant discovery (SNPs + Indels): GATK; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-.

472. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-+.

473. Handsaker RE, Korn JM, Nemesh J, McCarroll SA. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. Nat Genet. 2011;43(3):269-76.

474. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2003;1(1303).

475. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

476. Institute B. Picard 2020 [Available from: http://broadinstitute.github.io/picard/.

477. GATK. BaseRecalibrator: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360042477672-BaseRecalibrator.

478. GATK. AnalyzeCovariates: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360042911971-AnalyzeCovariates.

479. GATK. FisherStrand: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360036896952-FisherStrand.

480. GATK. *Tool Documentation Index: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360036896772--Tool-Documentation-Index.

481. GATK. QualByDepth: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360036896872-QualByDepth.

482. GATK. AS_ReadPosRankSumTest: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360037269371-AS-ReadPosRankSumTest.

483. GATK. Variant Quality Score Recalibration (VQSR): Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360035531612?id=1259.

484. GATK. Phred-scaled quality scores: Broad Institute; [Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360035531872-Phred-scaled-quality-scores.

485. Alwi ZB. The Use of SNPs in Pharmacogenomics Studies. Malays J Med Sci. 2005;12(2):4-12.

486. De Wit P, Pespeni MH, Palumbi SR. SNP genotyping and population genomics from expressed sequences - current advances and future possibilities. Mol Ecol. 2015;24(10):2310-23.

487. Gagnaire PA, Normandeau E, Côté C, Møller Hansen M, Bernatchez L. The genetic consequences of spatially varying selection in the panmictic American eel (Anguilla rostrata). Genetics. 2012;190(2):725-36.

488. Limborg MT, Helyar SJ, M. dB, Taylor M, Nielsen E, Ogden R, et al. Environmental selection on transcriptome-derived SNPs in a high gene

flow marine fish, the Atlantic herring (*Clupea harengus*). Molecular Ecology. 2012;21(15):3686-703.

489. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. High-resolution genotyping by amplicon melting analysis using LCGreen. Clin Chem. 2003;49(6 Pt 1):853-60.

490. Renaut S, Nolte AW, Rogers SM, Derome N, Bernatchez L. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along

gradients of ecological speciation in lake whitefish species

pairs (Coregonus spp.). Molecular Ecology. 2011;20:545-59.

491. O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med. 2013;5(3):28.

492. GATK. Genome Analysis Toolkit: Broad Institute; 2020 [Available from: https://gatk.broadinstitute.org/hc/en-us.

493. Sanggaard KW, Bechsgaard JS, Fang XD, Duan JJ, Dyrlund TF, Gupta V, et al. Spider genomes provide insight into composition and evolution of venom and silk (vol 5, 3765, 2014). Nat Commun. 2014;5.

494. Kim JS, Kim MJ, Kim HK, Vung NN, Kim I. Development of single nucleotide polymorphism markers specific to Apis mellifera (Hymenoptera: Apidae) line displaying high hygienic behavior against Varroa destructor, an ectoparasitic mite. J Asia-Pac Entomol. 2019;22(4):1031-9.

495. Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, Chikhi R, et al. Reference-free detection of isolated SNPs. Nucleic Acids Res. 2015;43(2):e11.

496. Quillery E, Quenez O, Peterlongo P, Plantard O. Development of genomic resources for the tick Ixodes ricinus: isolation and characterization of single nucleotide polymorphisms. Mol Ecol Resour. 2014;14(2):393-400.

497. Monzon JD, Atkinson EG, Henn BM, Benach JL. Population and Evolutionary Genomics of Amblyomma americanum, an Expanding Arthropod Disease Vector. Genome Biol Evol. 2016;8(5):1351-60.

498. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet. 2013;9(1):e1003215.

499. Lado P, Smith ML, Carstens BC, Klompen H. Population genetic structure and demographic history of the lone star tick, Amblyomma americanum (Ixodida: Ixodidae): New evidence supporting old records. Mol Ecol. 2020.

500. Lee HC, Lai K, Lorenc MT, Imelfort M, Duran C, Edwards D. Bioinformatics tools and databases for analysis of next-generation sequence data. Brief Funct Genomics. 2012;11(1):12-24.

501. Li RQ, Yu C, Li YR, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25(15):1966-7.

502. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008;18(11):1851-8.

503. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. Genome Res. 2001;11(10):1725-9.

504. Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat Rev Genet. 2018;19(5):269-85.

505. Glenn TC. Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011;11(5):759-69.

506. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-51.

507. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998;8(3):175-85.

508. Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, et al. Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res. 2008;18(5):763-70.

509. Li M, Nordborg M, Li LM. Adjust quality scores from alignment and improve sequencing accuracy. Nucleic Acids Research. 2004;32(17):5183-91.

510. Li RQ, Li YR, Fang XD, Yang HM, Wang J, Kristiansen K, et al. SNP detection for massively parallel whole-genome resequencing. Genome Res. 2009;19(6):1124-32.

511. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78-81.

512. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53-9.

513. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25(17):2283-5.

514. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008;452(7189):872-6.

515. Mokry M, Feitsma H, Nijman IJ, de Bruijn E, van der Zaag PJ, Guryev V, et al. Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. Nucleic Acids Res. 2010;38(10):e116.

516. Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20(2):273-80.

517. Hoberman R, Dias J, Ge B, Harmsen E, Mayhew M, Verlaan DJ, et al. A probabilistic approach for SNP discovery in high-throughput human resequencing data. Genome Res. 2009;19(9):1542-52.

518. Malhis N, Jones SJ. High quality SNP calling using Illumina data at shallow coverage. Bioinformatics. 2010;26(8):1029-35.

519. Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, et al. A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform. 2014;15(2):256-78.

520. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. Nature. 2011;470(7332):59-65.

521. Carbone I, Kohn L. Inferring process from pattern in fungal population genetics. Applied Mycology and Biotechnology. 2004;4:29-58.

522. Grünwald NJ, Everhart SE, Knaus BJ, Kamvar ZN. Best Practices for Population Genetic Analyses. Phytopathology. 2017;107(9):1000-10.

523. Hartl DL, Clark AG. Principles of Population Genetics. 3rd ed. Sunderland, MA: Sinauer Associates; 1997.

524. Phung TN, Huber CD, Lohmueller KE. Determining the Effect of Natural Selection on Linked Neutral Divergence across Species. PLoS Genet. 2016;12(8):e1006199.

525. Luikart G, Kardos M, Hand B, Rajora O, Aitken S, Hohenlohe P. Population Genomics: Advancing Understanding of Nature. In: Rajora O, editor. Population Genomics: Concepts, Approaches and Applications: Springer, Cham; 2018.

526. Hohenlohe PA, Phillips PC, Cresko WA. USING POPULATION GENOMICS TO DETECT SELECTION IN NATURAL POPULATIONS: KEY CONCEPTS AND METHODOLOGICAL CONSIDERATIONS. Int J Plant Sci. 2010;171(9):1059-71.

527. Black WC, Baer CF, Antolin MF, DuTeau NM. Population genomics: genome-wide sampling of insect populations. Annu Rev Entomol. 2001;46:441-69.

528. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4(12):981-94.

529. Schlötterer C. Hitchhiking mapping--functional genomics from the population genetics perspective. Trends Genet. 2003;19(1):32-8.

530. Stinchcombe JR, Hoekstra HE. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity (Edinb). 2008;100(2):158-70.
531. Maynard Smith J, Haigh J. The hitchhiking effect of a favourable gene. Genetics Research. 1974;23:23-35.
532. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. Nature Reviews Genetics. 2007;8(11):857-68.
533. Gillespie JH. The molecular nature of allelic diversity for two models of balancing selection. Theor Popul Biol. 1990;37(1):91-109.
534. Kreitman M, Hudson RR. Inferring the evolutionary histories of the Adh and Adh-dup loci in Drosophila melanogaster from patterns of polymorphism and divergence. Genetics. 1991;127(3):565-82.
535. Kaplan NL, Hudson RR, Langley CH. The Hitchhiking Effect Revisited. Genetics. 1989;123(4):887-99.
536. Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics. 2002;160(2):765-77.
537. Przeworski M. The signature of positive selection at randomly chosen loci. Genetics. 2002;160(3):1179-89.
538. McVean G. The structure of linkage disequilibrium around a selective sweep. Genetics. 2007;175(3):1395-406.
539. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The Hitchhiking Effect on the Site Frequency-Spectrum of DNA Polymorphisms. Genetics. 1995;140(2):783-96.
540. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics. 2000;155(3):1405-13.
541. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002;419(6909):832-7.
542. Voight BF, Kudaravalli S, Wen XQ, Pritchard JK. A map of recent positive selection in the human genome. Plos Biology. 2006;4(3):446-58.
543. Zeng K, Fu YX, Shi SH, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics. 2006;174(3):1431-9.

544. Cavalli-Sforza LL. Population structure and human evolution. Proc R Soc Lond B Biol Sci. 1966;164(995):362-79.

545. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 1973;74(1):175-95.

546. Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. Heredity (Edinb). 2009;103(4):285-98.

547. Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, et al. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol Ecol Resour. 2011;11 Suppl 1:123-36.

548. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol. 2014;29(1):51-63.

549. Hohenlohe PA, Hand BK, Andrews KR, Luikart G. Population genomics provide key insights in ecology and evolution. In: Rajora O, editor. Population Genomics: Springer, Cham; 2018. p. 483-510.

550. Martins NE, Faria VG, Nolte V, Schlötterer C, Teixeira L, Sucena É, et al. Host adaptation to viruses relies on few genes with different cross-resistance properties. Proc Natl Acad Sci U S A. 2014;111(16):5938-43.

551. Kofler R, Betancourt A, Schlötterer C. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. PLoS Genetics. 2012;8(1):e1002487.

552. Kim S, Lohmueller K, Albrechtsen A, Li Y, Korneliussen T, Tian G, et al. Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics. 2011;12:231.

553. Tsagkogeorga G, Cahais V, Galtier N. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate Ciona intestinalis. Genome Biol Evol. 2012;4(8):740-9.

554. Gayral P, Melo-Ferreira J, Glemin S, Bierene N, Carneiro M, Nahbolz B, et al. Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. PLoS Genetics. 2013;9(4):e1003457.

555. Han E, Sinsheimer J, Novembre J. Characterizing bias in population genetic inferences from low-coverage sequencing data. Molecular Biology and Evolution. 2014;31(3):723-35.

556. Han B, Kang EY, Raychaudhuri S, de Bakker PIW, Eskin E. Fast pairwise IBD association testing in genome-wide association studies. Bioinformatics. 2014;30:206-13.

557. Nielsen R, Koreneliussen T, Albrectsen A, Li Y, Wang J. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS ONE. 2012;7:e37558.

558. Brumfield R, Beerli P, Nickerson D, Edwards S. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution. 2003;18(5):249-56.

559. Morin P, McCarthy M. Highly accurate SNP genotyping from historical and low-quality samples. Molecular Ecology Notes. 2007;7:937-46.

560. Primmer CR, Borge T, Lindell J, Saetre GP. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Mol Ecol. 2002;11(3):603-12.

561. Aitken N, Smith S, Schwarz C, Morin PA. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol Ecol. 2004;13(6):1423-31.

562. Elfstrom C, Smith C, Seeb J. Thirty-two single nucleotide polymorphism markers for high-throughput genotyping of sockeye salmon. 2006;6(4):1255-9.

563. Kuhner MK, Beerli P, Yamato J, Felsenstein J. Usefulness of single nucleotide polymorphism data for estimating population parameters. Genetics. 2000;156(1):439-47.

564. Glaubitz JC, Rhodes OE, Dewoody JA. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Mol Ecol. 2003;12(4):1039-47.

565. Smith S, Aitken N, Schwarz C, Morin P. Characterisation of 15 single nucleotide polymorphism markers for chimpanzees (*Pan troglodytes*). Molecular Ecology Notes. 2004;4:348-51.

566. Seddon JM, Parker HG, Ostrander EA, Ellegren H. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. Mol Ecol. 2005;14(2):503-11.

567. Kohn MH, Murphy WJ, Ostrander EA, Wayne RK. Genomics and conservation genetics. Trends Ecol Evol. 2006;21(11):629-37.

568. Nielsen R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics. 2000;154(2):931-42.

569. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. Electrophoresis. 1999;20(8):1682-96.
570. Hedrick PW. A standardized genetic differentiation measure. Evolution. 2005;59(8):1633-8.
571. Kalinowski ST. How many alleles per locus should be used to estimate genetic distances? Heredity (Edinb). 2002;88(1):62-5.
572. Ryman N, Palm S, André C, Carvalho GR, Dahlgren TG, Jorde PE, et al. Power for detecting genetic divergence: differences between statistical methods and marker loci. Mol Ecol. 2006;15(8):2031-45.
573. Ryman N, Palm S. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Molecular Ecology Notes. 2006;6:600-2.
574. Ryman N. CHIFISH: a computer program testing for genetic heterogeneity at multiple loci using chi-square and Fishers's exact test. Molecular Ecology Notes. 2006;6:285-7.
575. Smith MJ, Pascal CE, Grauvogel Z, Habicht C, Seeb JE, Seeb LW. Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales. Mol Ecol Resour. 2011;11 Suppl 1:268-77.
576. Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Mol Ecol Resour. 2011;11 Suppl 1:254-67.
577. Karlsson S, Moen T, Lien S, Glover KA, Hindar K. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol Ecol Resour. 2011;11 Suppl 1:247-53.
578. Geraldes A, Pang J, Thiessen N, Cezard T, Moore R, Zhao Y, et al. SNP discovery in black cottonwood (Populus trichocarpa) by population transcriptome resequencing. Mol Ecol Resour. 2011;11 Suppl 1:81-92.
579. Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Resour. 2011;11 Suppl 1:117-22.
580. Martinez DA, Nelson MA. The next generation becomes the now generation. PLoS Genetics. 2010;6(4):e100096.

581. Morin PA, Pease VL, Hancock BL, Roberston KM, Antolik CW, Huebinger R. Characterization of 42 single nucleotide polymorphism (SNP) markers for the bowhead whale (*Balaena mysticetus*) for use in discriminating populations. Marine Mammal Science. 2010;26:716-32.
582. Campbell NR, Narum SR. Development of 54 novel single-nucleotide polymorphism (SNP) assays for sockeye and coho salmon and assessment of available SNPs to differentiate stocks within the Columbia River. Mol Ecol Resour. 2011;11 Suppl 1:20-30.
583. Dai Z, Papp AC, Wang D, Hampel H, Sadee W. Genotyping panel for assessing response to cancer chemotherapy. BMC Med Genomics. 2008;1:24.
584. Nielsen EE, Cariani A, Mac Aoidh E, Maes GE, Milano I, Ogden R, et al. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat Commun. 2012;3:851.
585. Banks MA, Eichert W, Olsen JB. Which genetic loci have greater population assignment power? Bioinformatics. 2003;19(11):1436-8.
586. Bromaghin JF. BELS: backward elimination locus selection for studies of mixture composition or individual assignment. Molecular Ecology Resources. 2008;8:568-71.
587. Glover KA, Hansen MM, Lien S, Als TD, Høyheim B, Skaala O. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genet. 2010;11:2.
588. Anderson EC. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Mol Ecol Resour. 2010;10(4):701-10.
589. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011;12(9):628-40.
590. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012;20(5):490-7.
591. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nature Reviews Genetics. 2011;12(11):745-55.
592. Katsavou E, Vlogiannitis S, Karp-Tatham E, Blake DP, Ilias A, Strube C, et al. Identification and geographical distribution of pyrethroid resistance mutations in the poultry red mite Dermanyssus gallinae. Pest Manag Sci. 2020;76(1):125-33.

593. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Mol Biol Evol. 2017;34(3299-3302).

594. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics. 2000;16(9):847-8.

595. Haubold B, Travisano M, Rainey PB, Hudson RR. Detecting linkage disequilibrium in bacterial populations. Genetics. 1998;150(4):1341-8.

596. Shen TH, Carlson CS, Tarczy-Hornoch P. SNPit: a federated data integration system for the purpose of functional SNP annotation. Comput Methods Programs Biomed. 2009;95(2):181-9.

597. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics. 2009;25(5):655-61.

598. Wang P, Dai M, Xuan W, McEachin RC, Jackson AU, Scott LJ, et al. SNP Function Portal: a web database for exploring the function implication of SNP alleles. Bioinformatics. 2006;22(14):e523-9.

599. Riva A, Kohane IS. SNPper: retrieval and analysis of human SNPs. Bioinformatics. 2002;18(12):1681-5.

600. Liu CK, Chen YH, Tang CY, Chang SC, Lin YJ, Tsai MF, et al. Functional analysis of novel SNPs and mutations in human and mouse genomes. BMC Bioinformatics. 2008;9 Suppl 12:S10.

601. Goodswen SJ, Gondro C, Watison-Haigh NS, Kadarmideen HN. FunctSNP: an R package to link SNPs to functional knowledge and dbAutoMaker: a suite of Perl scripts to build SNP databases. BMC Bioinformatics. 2010;11:311.

602. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

603. Doran AG, Creevey CJ. Snpdat: easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms. BMC Bioinformatics. 2013;14:45.

604. Bohonak AJ. Dispersal, gene flow, and population structure. Q Rev Biol. 1999;74(1):21-45.

605. Levy F, Neal CL. Spatial and temporal genetic structure in chloroplast and allozyme markers in phacelia dubia implicate genetic drift. Heredity (Edinb). 1999;82 Pt 4:422-31.

606. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-85.

607. Mitton JB, Koehn RK, Prout T. Population genetics of marine pelecypods. III. Epistasis between functionally related isoenzymes of M*ytilus edulis*. Genetics. 1973;73:487-96.

608. Nei M, Li WH. Linkage disequilibrium in subdivided populations. Genetics. 1973;75(1):213-9.

609. Li WH, Nei M. Stable linkage disequilibrium without epistasis in subdivided populations. Theor Popul Biol. 1974;6(2):173-83.

610. Slatkin M. Gene flow and selection in a 2-locus system. Genetics. 1975;81:787-802.

611. Wang J, Caballero A, Hill W. The effect of linkage disequilibrium and deviation from Hardy–Weinberg proportions on the changes in genetic variance with bottlenecking. Heredity. 1998;81:174-86.

612. Chen X, Sullivan PF. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenomics J. 2003;3(2):77-96.

613. Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A. Detecting selective sweeps from pooled next-generation sequencing samples. Mol Biol Evol. 2012;29(9):2177-86.

614. Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, et al. Ecological genomics of Anopheles gambiae along a latitudinal cline: a population-resequencing approach. Genetics. 2012;190(4):1417-32.

615. Bajda S, Riga M, Wybouw N, Papadaki S, Ouranou E, Fotoukkiaii SM, et al. Fitness costs of key point mutations that underlie acaricide target-site resistance in the two-spotted spider mite Tetranychus urticae. Evolutionary Applications. 2018;11(9):1540-53.

616. Sparks TC. IRAC: Mode of action classification and insecticide resistance management. Abstracts of Papers of the American Chemical Society. 2014;248.

617. Van Leeuwen T, Vontas J, Tsagkarakou A, Dermauw W, Tirry L. Acaricide resistance mechanisms in the two-spotted spider mite Tetranychus urticae and other important Acari: a review. Insect Biochem Mol Biol. 2010;40(8):563-72.

618. Feyereisen R, Dermauw W, Van Leeuwen T. Genotype to phenotype, the molecular and physiological dimensions of resistance in arthropods. Pestic Biochem Physiol. 2015;121:61-77.

619. Riga M, Bajda S, Themistokleous C, Papadaki S, Palzewicz M, Dermauw W, et al. The relative contribution of target-site mutations in complex acaricide resistant phenotypes as assessed by marker assisted backcrossing in Tetranychus urticae. Sci Rep-Uk. 2017;7.

620. Sparks TC, Nauen R. IRAC: Mode of action classification and insecticide resistance management. Pestic Biochem Phys. 2015;121:122-8.

621. Sparks TC. Insecticide discovery: An evaluation and analysis. Pestic Biochem Phys. 2013;107(1):8-17.

622. Van Leeuwen T, Dermauw W. The Molecular Evolution of Xenobiotic Metabolism and Resistance in Chelicerate Mites. Annu Rev Entomol. 2016;61:475-98.

623. Raymond M, Heckel DG, Scott JG. Interactions between Pesticide Genes - Model and Experiment. Genetics. 1989;123(3):543-51.

624. Bohannan BJM, Travisano M, Lenski RE. Epistatic interactions can lower the cost of resistance to multiple consumers. Evolution. 1999;53(1):292-5.

625. Williams T, Mwangi T, Wambua S, Uyoga S, Weatherall D, Gupta S, et al. Negative epistasis between the malaria-protective effects of alpha (+) thalassaemia and the sickle cell trait [MIM-TW-395505]. Acta Trop. 2005;95:S91-S.

626. Moore JH, Williams SM. Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. Bioessays. 2005;27(6):637-46.

627. Zhang YX, Meng XK, Yang YX, Li H, Wang X, Yang BJ, et al. Synergistic and compensatory effects of two point mutations conferring target-site resistance to fipronil in the insect GABA receptor RDL. Sci Rep-Uk. 2016;6.

628. Hardstone MC, Scott JG. A review of the interactions between multiple insecticide resistance loci. Pestic Biochem Phys. 2010;97(2):123-8.

629. Shi MA, Lougarre A, Alies C, Frémaux I, Tang ZH, Stojan J, et al. Acetylcholinesterase alterations reveal the fitness cost of mutations conferring insecticide resistance. BMC Evol Biol. 2004;4:5.
630. Liu NN, Pridgeon JW. Metabolic detoxication and the kdr mutation in pyrethroid resistant house flies, M*usca domestica* (L.).
Pesticide B**iochemistry** and P**hysiology**. 2002;73:157-63.
631. Peyronnet O, Pichon Y, Carton Y, Delorme R. Genetic-analysis of deltamethrin resistance in laboratory-selected strains of D*rosophila melanogaster* Meig.
Pesticide Biochemistry and Physiology. 1994;50:207-18.
632. McEnroe WD, Naegele JA. The coadaptive process in an organophosphorus-resistant strain of the two-spotted spider mite, T*etranychus urticae*. Annals of the Entomological Society of America. 1968;61:1055-9.
633. Georghiou GP. Genetics of resistance to insecticide in houseflies and mosquitoes. Experimental Parasitology. 1969;26:244-55.
634. Mckenzie JA, Whitten MJ, Adena MA. The effect of genetic background on the fitness of diazinon resistance genotypes of the Australian Sheep Blowfly, L*ucilia cuprina*. Heredity. 1982;49:1-9.
635. Abbas RZ, Colwell DD, Iqbal Z, Khan A. Acaricidal drug resistance in poultry red mite (Dermanyssus gallinae) and approaches to its management. World Poultry Science Association. 2014;70:113-24.
636. Collison CH, Danka RG, Kennell DR. An evaluation of permethrin, carbaryl, and amitraz for the control of northern fowl mites on caged chickens. Poult Sci. 1981;60(8):1812-7.
637. Ware GW. The Pesticide Book. California: Thomson Publications; 2000.
638. Narahasi T. Mode of action of pyrethroids. Bulletin of World Health Organisation. 1971;44:337-45.
639. Soderlund DM, Knipple DC. The molecular biology of knockdown resistance to pyrethroid insecticides. Insect Biochem Molec. 2003;33(6):563-77.
640. Khambay B, Jewess P. Pyrethroids. In: Iatrou K, Gilbert LI, Gill SS, editors. Comprehensive Molecular Insect Science. 6. Oxford, UK: Elsevier; 2005. p. 1-29.
641. Zhu F, Parthasarathy R, Bai H, Woithe K, Kaussmann M, Nauen R, et al. A brain-specific cytochrome P450 responsible for the majority of deltamethrin resistance in

the QTC279 strain of Tribolium castaneum. Proc Natl Acad Sci U S A. 2010;107(19):8557-62.

642. Enayati AA, Asgarian F, Amouei A, Sharif M, Mortazavi H, Boujhmehrani H, et al. Pyrethroid insecticide resistance in *Rhipicephalus bursa* (Acari, Ixodidae). Pesticide Biochemistry and Physiology. 2010;97:125-8.

643. Zuo YY, Peng X, Wang K, Lin FF, Li YT, Chen MH. Expression patterns, mutation detection and RNA interference of Rhopalosiphum padi voltage-gated sodium channel genes. Sci Rep-Uk. 2016;6.

644. Wang R, Liu Z, Dong K, Elzen P, Pettis J, Huang Z. Association of novel mutations in a sodium channel gene with fluvalinate resistance in the mite, V*arroa destructor*. Journal of Apicultural Research. 2002;41(1-2):17-25.

645. Ay R, Gurkan MO. Resistance to bifenthrin and resistance mechanisms of different strains of the two-spotted spider mite (Tetranychus urticae) from Turkey. Phytoparasitica. 2005;33(3):237-44.

646. Van Leeuwen T, Van Pottelberge S, Tirry L. Comparative acaricide susceptibility and detoxifying enzyme activities in field-collected resistant and susceptible strains of Tetranychus urticae. Pest Management Science. 2005;61(5):499-507.

647. Van Leeuwen T, Tirry L. Esterase-mediated bifenthrin resistance in a multiresistant strain of the two-spotted spider mite, Tetranychus urticae. Pest Management Science. 2007;63(2):150-6.

648. Feng YN, Zhao S, Sun W, Li M, Lu WC, He L. The sodium channel gene in Tetranychus cinnabarinus (Boisduval): identification and expression analysis of a mutation associated with pyrethroid resistance. Pest Manag Sci. 2011;67(8):904-12.

649. Dong K, Du YZ, Rinkevich F, Nomura Y, Xu P, Wang LX, et al. Molecular biology of insect sodium channels and pyrethroid resistance. Insect Biochem Molec. 2014;50:1-17.

650. Davies TGE, O'Reilly AO, Field LM, Wallace BA, Williamson MS. Knockdown resistance to DDT and pyrethroids: from target-site mutations to molecular modelling. Pest Management Science. 2008;64(11):1126-30.

651. Tsagkarakou A, Van Leeuwen T, Khajehali J, Ilias A, Grispou M, Williamson MS, et al. Identification of pyrethroid resistance associated mutations in the para sodium channel of the two-spotted spider mite Tetranychus urticae (Acari: Tetranychidae). Insect Mol Biol. 2009;18(5):583-93.

652. He H, Chen AC, Davey RB, Ivie GW, George JE. Identification of a point mutation in the para-type sodium channel gene from a pyrethroid-resistant cattle tick. Biochem Biophys Res Commun. 1999;261(3):558-61.

653. Tan J, Liu Z, Wang R, Huang ZY, Chen AC, Gurevitz M, et al. Identification of amino acid residues in the insect sodium channel critical for pyrethroid binding. Mol Pharmacol. 2005;67(2):513-22.

654. Nyoni BN, Gorman K, Mzilahowa T, Williamson MS, Navajas M, Field LM, et al. Pyrethroid resistance in the tomato red spider mite, Tetranychus evansi, is associated with mutation of the para-type sodium channel. Pest Manag Sci. 2011;67(8):891-7.

655. Du Y, Garden DP, Wang L, Zhorov BS, Dong K. Identification of new batrachotoxin-sensing residues in segment IIIS6 of the sodium channel. J Biol Chem. 2011;286(15):13151-60.

656. Wrzesińska B, Czerwoniec A, Wieczorek P, Węgorek P, Zamojska J, Obrępalska-Stęplowska A. A survey of pyrethroid-resistant populations of Meligethes aeneus F. in Poland indicates the incidence of numerous substitutions in the pyrethroid target site of voltage-sensitive sodium channels in individual beetles. Insect Mol Biol. 2014;23(5):682-93.

657. Usherwood PN, Davies TG, Mellor IR, O'Reilly AO, Peng F, Vais H, et al. Mutations in DIIS5 and the DIIS4-S5 linker of Drosophila melanogaster sodium channel define binding domains for pyrethroids and DDT. FEBS Lett. 2007;581(28):5485-92.

658. Vais H, Williamson MS, Devonshire AL, Usherwood PN. The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. Pest Manag Sci. 2001;57(10):877-88.

659. Moyes CL, Vontas J, Martins AJ, Ng LC, Koou SY, Dusfour I, et al. Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. PLoS Negl Trop Dis. 2017;11(7):e0005625.

660. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dokel S, Ralser M, et al. Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics. 2014;15:675.

661. de la Fuente J. Vaccines for vector control: exciting possibilities for the future. Vet J. 2012;194(2):139-40.

662. Otranto D. Arthropod-borne pathogens of dogs and cats: From pathways and times of transmission to disease control. Vet Parasitol. 2018;251:68-77.

663. de la Fuente J, Villar M, Contreras M, Moreno-Cid JA, Merino O, Perez de la Lastra JM, et al. Prospects for vaccination against the ticks of pets and the potential impact on pathogen transmission. Vet Parasitol. 2015;208(1-2):26-9.

664. Vonlaufen N, Kanzok SM, Wek RC, Sullivan WJ, Jr. Stress response pathways in protozoan parasites. Cell Microbiol. 2008;10(12):2387-99.

665. Zarnbrano-Villa S, Rosales-Borjas D, Carrero JC, Ortiz-Ortiz L. How protozoan parasites evade the immune response. Trends Parasitol. 2002;18(6):272-8.

666. Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat Rev Microbiol. 2009;7(7):493-503.

667. Karls AC, Perkins-Balding D. Antigenic Variation. In: Maloy S, Hughes K, editors. Brenner's Encyclopedia of Genetics. 2nd ed. London: Academic Press; 2013. p. 145-7.

668. Morris RV, Shoemaker CB, David JR, Lanzaro GC, Titus RG. Sandfly maxadilan exacerbates infection with Leishmania major and vaccinating against it protects against L. major infection. J Immunol. 2001;167(9):5226-30.

669. Valenzuela JG, Belkaid Y, Garfield MK, Mendez S, Kamhawi S, Rowton ED, et al. Toward a defined anti-Leishmania vaccine targeting vector antigens: characterization of a protective salivary protein. J Exp Med. 2001;194(3):331-42.

670. Mejia JS, Bishop JV, Titus RG. Is it possible to develop pan-arthropod vaccines? Trends Parasitol. 2006;22(8):367-70.

671. Arca B, Lombardo F, Valenzuela JG, Francischetti IM, Marinotti O, Coluzzi M, et al. An updated catalogue of salivary gland transcripts in the adult female mosquito, Anopheles gambiae. J Exp Biol. 2005;208(Pt 20):3971-86.

672. Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Mather TN, Ribeiro JM. Exploring the sialome of the tick Ixodes scapularis. J Exp Biol. 2002;205(Pt 18):2843-64.

673. Milleron RS, Mutebi JP, Valle S, Montoya A, Yin H, Soong L, et al. Antigenic diversity in maxadilan, a salivary protein from the sand fly vector of American visceral leishmaniasis. Am J Trop Med Hyg. 2004;70(3):286-93.

674. Milleron RS, Ribeiro JM, Elnaime D, Soong L, Lanzaro GC. Negative effect of antibodies against maxadilan on the fitness of the sand fly vector of American visceral leishmaniasis. Am J Trop Med Hyg. 2004;70(3):278-85.

675. Weedall GD, Conway DJ. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. Trends Parasitol. 2010;26(7):363-9.

676. Barry AE, Schultz L, Buckee CO, Reeder JC. Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, Plasmodium falciparum. PLoS One. 2009;4(12):e8497.
677. Nóbrega de Sousa T, Carvalho LH, Alves de Brito CF. Worldwide genetic variability of the Duffy binding protein: insights into Plasmodium vivax vaccine development. PLoS One. 2011;6(8):e22944.
678. Takala SL, Coulibaly D, Thera MA, Batchelor AH, Cummings MP, Escalante AA, et al. Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. Sci Transl Med. 2009;1(2):2ra5.
679. Arnott A, Mueller I, Ramsland PA, Siba PM, Reeder JC, Barry AE. Global Population Structure of the Genes Encoding the Malaria Vaccine Candidate, Plasmodium vivax Apical Membrane Antigen 1 (PvAMA1). PLoS Negl Trop Dis. 2013;7(10):e2506.
680. Conway DJ, Cavanagh DR, Tanabe K, Roper C, Mikes ZS, Sakihama N, et al. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nat Med. 2000;6(6):689-92.
681. Duan J, Mu J, Thera MA, Joy D, Kosakovsky Pond SL, Diemert D, et al. Population structure of the genes encoding the polymorphic Plasmodium falciparum apical membrane antigen 1: implications for vaccine design. Proc Natl Acad Sci U S A. 2008;105(22):7857-62.
682. Barry AE, Arnott A. Strategies for designing and monitoring malaria vaccines targeting diverse antigens. Front Immunol. 2014;5:359.
683. Chan ER, Menard D, David PH, Ratsimbasoa A, Kim S, Chim P, et al. Whole genome sequencing of field isolates provides robust characterization of genetic diversity in Plasmodium vivax. PLoS Negl Trop Dis. 2012;6(9):e1811.
684. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing Patterns of Selection and Geospatial Genetic Diversity within Two Leading Plasmodium vivax Candidate Vaccine Antigens. Plos Neglect Trop D. 2014;8(4).
685. Ouattara A, Takala-Harrison S, Thera MA, Coulibaly D, Niangaly A, Saye R, et al. Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications. J Infect Dis. 2013;207(3):511-9.

686. Takala SL, Coulibaly D, Thera MA, Batchelor AH, Cummings MP, Escalante AA, et al. Extreme Polymorphism in a Vaccine Antigen and Risk of Clinical Malaria: Implications for Vaccine Development. Science Translational Medicine. 2009;1(2).

687. Dutta S, Dlugosz LS, Drew DR, Ge XP, Ababacar D, Rovira YI, et al. Overcoming Antigenic Diversity by Enhancing the Immunogenicity of Conserved Epitopes on the Malaria Vaccine Candidate Apical Membrane Antigen-1. Plos Pathogens. 2013;9(12).

688. Bai T, Becker M, Gupta A, Strike P, Murphy VJ, Anders RF, et al. Structure of AMA1 from Plasmodium falciparum reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. Proc Natl Acad Sci U S A. 2005;102(36):12736-41.

689. Coley AM, Parisi K, Masciantonio R, Hoeck J, Casey JL, Murphy VJ, et al. The most polymorphic residue on Plasmodium falciparum apical membrane antigen 1 determines binding of an invasion-inhibitory antibody. Infect Immun. 2006;74(5):2628-36.

690. Tetteh KK, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, Marsh K, et al. Prospective identification of malaria parasite genes under balancing selection. PLoS One. 2009;4(5):e5568.

691. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in Plasmodium falciparum. Mol Biol Evol. 2010;27(10):2344-51.

692. Genton B, Betuela I, Felger I, Al-Yaman F, Anders RF, Saul A, et al. A recombinant blood-stage malaria vaccine reduces Plasmodium falciparum density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. J Infect Dis. 2002;185(6):820-7.

693. Comas D, Piulachs M, Belles X. Vitellogenin of *Blattella germanica* (L.) (Dictyoptera,Blattellidae): Nucleotide Sequence of the cDNA and Analysisof the Protein Primary Structure. Arch Insect Biochem. 2000;45:1-11.

694. Zhang T, Zhang G, Zeng F, Mao J, Liang H, Liu F. Molecular Cloning of the Vitellogenin Gene and the Effects of Vitellogenin Protein Expression on the Physiology of Harmonia axyridis (Coleoptera: Coccinellidae). Sci Rep. 2017;7(1):13926.

695. Hagedorn HH, Maddison DR, Tu ZJ. The evolution of vitellogenins, cyclorrhaphan yolk proteins and related molecules. Adv Insect Physiol. 1998;27:335-84.

696. Sappington TW, Raikhel AS. Molecular characteristics of insect vitellogenins and vitellogenin receptors. Insect Biochem Molec. 1998;28(5-6):277-300.

697. Boldbaatar D, Umemiya-Shirafuji R, Liao M, Tanaka T, Xuan X, Fujisaki K. Multiple vitellogenins from the Haemaphysalis longicornis tick are crucial for ovarian development. J Insect Physiol. 2010;56(11):1587-98.

698. Wang SY, Smith DE, Williams DL. Purification of avian vitellogenin III: comparison with vitellogenins I and II. Biochemistry-Us. 1983;22:6206-12.

699. Hirai M, Watanabe D, Kiyota A, Chinzei Y. Nucleotide sequence of vitellogenin mRNA in the bean bug, Riptortus clavatus: analysis of processing in the fat body and ovary. Insect Biochem Mol Biol. 1998;28(8):537-47.

700. Serrano-Pinto V, Landais I, Ogliastro MH, Gutiérrez-Ayala M, Mejía-Ruíz H, Villarreal-Colmenares H, et al. Vitellogenin mRNA expression in Cherax quadricarinatus during secondary vitellogenic at first maturation females. Mol Reprod Dev. 2004;69(1):17-21.

701. Yang WJ, Ohira T, Tsutsui N, Subramoniam T, Huong DT, Aida K, et al. Determination of amino acid sequence and site of mRNA expression of four vitellins in the giant fresh water prawn, Macrobrachium rosenbergii. Journal of Experimental Zoology. 2000;287(413-422).

702. Donohue KV, Khalil SMS, Mitchell RD, Sonenshine DE, Roe RM. Molecular characterization of the major hemelipoglycoprotein in ixodid ticks. Insect molecular biology. 2008;17(3):197-208.

703. Zmora N, Trant J, Chan SM, Chung JS. Vitellogenin and its mRNA during ovarian development in the female blue crub, Callinectes sapidus: gene expression, synthesis, transport and cleavage. Biology of Reproduction. 2007;77:138-46.

704. Avarre JC, Michelis R, Tietz A, Lubzens E. Relationship between vitellogenin and vitellin in a marine shrimp (Penaeus semisulcatus) and molecular characterization of vitellogenin complementary DNAs. Biology of Reproduction. 2003;69(1):355-64.

705. Thomson DM, Khalil SM, Jeffers LA, Ananthapadmanaban U, Sonenshine DE, Mitchell RD, et al. In vivo role of 20-hydroxyecdysone in the regulation of the

vitellogenin mRNA and eggs development in the American dog tick D*ermacentor variabilis* Insect Physiology. 2005;51:1105 - 16.
706. Gudderra NP, Neese PA, Sonenshine DE, Apperson CS, Roe RM. Developmental profile, isolation, and biochemical characterization of a novel lipoglycoheme-carrier protein from the American dog tick, Dermacentor variabilis (Acari: Ixodidae) and observations on a similar protein in the soft tick, Ornithodoros parkeri (Acari: Argasidae). Insect Biochem Mol Biol. 2001;31(4-5):299-311.
707. Sullivan CD, Rosell RC, Coons LB. Partial characterization of vitellogenin from the ixodid Dermacentor variabilis: preliminary results. In: Needham GR, Mitchell R, Horn DJ, Welbourn WC, editors. Acarology XI. 2. Columbus, Ohio: Ohio Biological Survey 1999.
708. Thompson DM, Khalil SM, Jeffers LA, Sonenshine DE, Mitchell RD, Osgood CJ, et al. Sequence and the development and tissue-specific regulation of the first complete vitellogenin messenger RNA from ticks responsible for heme sequestration. Insect Biochem Molec. 2007;37(363-374).
709. Taylor D, Chinzei Y, Miura K, Ando K. Vitellogenin synthesis, processing and hormonal regulation in the tick, Ornithodoros parkeri (Acari:Argasidae). Insect Biochem Mol Biol. 1991;21:723-33.
710. Tellam RL, Kemp D, Riding G, Briscoe S, Smith D, Sharp P, et al. Reduced oviposition of Boophilus microplus feeding on sheep vaccinated with vitellin. Vet Parasitol. 2002;103(1-2):141-56.
711. James AM, Oliver JH. Localization of vitellogenin production in the black legged tick, Ixodes scapularis (Acari: Ixodidae). Invertebrate Reproduction and Development. 1999;35:81-7.
712. Morandin C, Havukainen H, Kulmuni J, Dhaygude K, Trontii K, Helantera H. Not only for egg yolk – functional and evolutionary insights from expression, selection, and structural analyses of Formica ant vitellogenins. Mol Biol Evol. 2014;31:2181-93.
713. Mitchell RD, 3rd, Ross E, Osgood C, Sonenshine DE, Donohue KV, Khalil SM, et al. Molecular characterization, tissue-specific expression and RNAi knockdown of the first vitellogenin receptor from a tick. Insect Biochem Mol Biol. 2007;37(4):375-88.
714. Smith AD, Reuben Kaufman W. Molecular characterization of the vitellogenin receptor from the tick, Amblyomma hebraeum (Acari: Ixodidae). Insect Biochem Mol Biol. 2013;43(12):1133-41.

715. Benes P, Vetvicka V, Fusek M. Cathepsin D-Many functions of one aspartic protease. Crit Rev Oncol Hemat. 2008;68(1):12-28.
716. Erickson AH, Conner GE, Blobel G. Biosynthesis of a Lysosomal-Enzyme - Partial Structure of 2 Transient and Functionally Distinct Nh2-Terminal Sequences in Cathepsin-D. Journal of Biological Chemistry. 1981;256(21):1224-31.
717. Conner GE, Richo G. Isolation and Characterization of a Stable Activation Intermediate of the Lysosomal Aspartyl Protease Cathepsin-D. Biochemistry-Us. 1992;31(4):1142-7.
718. Gieselmann V, Hasilik A, von Figura K. Processing of human cathepsin D in lysosomes in vitro. J Biol Chem. 1985;260(5):3215-20.
719. Dunn BM, Scarborough PE, Lowther WT, Rao-Naik C. Comparison of the active site specificity of the aspartic proteinases based on a systematic series of peptide substrates. Adv Exp Med Biol. 1995;362:1-9.
720. Dunn BM, Hung S. The two sides of enzyme-substrate specificity: lessons from the aspartic proteinases. Biochim Biophys Acta. 2000;1477(1-2):231-40.
721. Baechle D, Flad T, Cansier A, Steffen H, Schittek B, Tolson J, et al. Cathepsin D is present in human eccrine sweat and involved in the postsecretory processing of the antimicrobial peptide DCD-1L. Journal of Biological Chemistry. 2006;281(9):5406-15.
722. Hakala JK, Oksjoki R, Laine P, Du H, Grabowski GA, Kovanen PT, et al. Lysosomal enzymes are released from cultured human macrophages, hydrolyze LDL in vitro, and are present extracellularly in human atherosclerotic lesions. Arterioscler Thromb Vasc Biol. 2003;23(8):1430-6.
723. Bankowska A, Gacko M, Chyczewska E, Worowska A. Biological and diagnostic role of cathepsin D. Rocz Akad Med Bialymst. 1997;42 Suppl 1:79-85.
724. Takei Y, Higashira H, Yamamoto T, Hayashi K. Mitogenic activity toward human breast cancer cell line MCF-7 of two bFGFs purified from sera of breast cancer patients: co-operative role of cathepsin D. Breast Cancer Res Treat. 1997;43(1):53-63.
725. Morikawa W, Yamamoto K, Ishikawa S, Takemoto S, Ono M, Fukushi J, et al. Angiostatin generation by cathepsin D secreted by human prostate carcinoma cells. J Biol Chem. 2000;275(49):38912-20.
726. Piwnica D, Touraine P, Struman I, Tabruyn S, Bolbach G, Clapp C, et al. Cathepsin D processes human prolactin into multiple 16K-like N-terminal fragments:

Study of their antiangiogenic properties and physiological relevance. Mol Endocrinol. 2004;18(10):2522-42.
727. Lkhider M, Castino R, Bouguyon E, Isidoro C, Ollivier-Bousquet M. Cathepsin D released by lactating rat mammary epithelial cells is involved in prolactin cleavage under physiological conditions. Journal of Cell Science. 2004;117(21):5155-64.
728. Ferreras M, Felbor U, Lenhard T, Olsen BR, Delaisse JM. Generation and degradation of human endostatin proteins by various proteinases. Febs Lett. 2000;486(3):247-51.
729. Baumgrass R, Williamson MK, Price PA. Identification of peptide fragments generated by digestion of bovine and human osteocalcin with the lysosomal proteinases cathepsin B, D, L, H, and S. J Bone Miner Res. 1997;12(3):447-55.
730. Woessner JF, Jr., Shamberger RJ, Jr. Purification and properties of cathepsin D from bovine utrus. J Biol Chem. 1971;246(7):1951-60.
731. Sajid M, McKerrow JH. Cysteine proteases of parasitic organisms. Mol Biochem Parasit. 2002;120(1):1-21.
732. Sojka D, Hartmann D, Bartosova-Sojkova P, Dvorak J. Parasite Cathepsin D-Like Peptidases and Their Relevance as Therapeutic Targets. Trends Parasitol. 2016;32(9):708-23.
733. Mahanti M, Bhakat S, Nilsson UJ, Soderhjelm P. Flap Dynamics in Aspartic Proteases: A Computational Perspective. Chem Biol Drug Des. 2016;88(2):159-77.
734. Mulenga A, Sugimoto C, Ingram G, Ohashi K, Onuma M. Molecular cloning of two Haemaphysalis longicornis cathepsin L-like cysteine proteinase genes. J Vet Med Sci. 1999;61(5):497-502.
735. Renard G, Garcia JF, Cardoso FC, Richter MF, Sakanari JA, Ozaki LS, et al. Cloning and functional expression of a Boophilus microplus cathepsin L-like enzyme. Insect Biochem Mol Biol. 2000;30(11):1017-26.
736. Seixas A, Estrela AB, Ceolato JC, Pontes EG, Lara F, Gondim KC, et al. Localization and function of Rhipicephalus (Boophilus) microplus vitellin-degrading cysteine endopeptidase. Parasitology. 2010;137(12):1819-31.
737. Seixas A, Dos Santos PC, Velloso FF, Da Silva Vaz I, Jr., Masuda A, Horn F, et al. A Boophilus microplus vitellin-degrading cysteine endopeptidase. Parasitology. 2003;126(Pt 2):155-63.

738. Yamaji K, Tsuji N, Miyoshi T, Islam MK, Hatta T, Alim MA, et al. Hemoglobinase activity of a cysteine protease from the ixodid tick Haemaphysalis longicornis. Parasitol Int. 2009;58(3):232-7.

739. Estrela AB, Seixas A, Teixeira Vde O, Pinto AF, Termignoni C. Vitellin- and hemoglobin-digesting enzymes in Rhipicephalus (Boophilus) microplus larvae and females. Comp Biochem Physiol B Biochem Mol Biol. 2010;157(4):326-35.

740. Mans BJ. Evolution of Vertebrate Hemostatic and Inflammatory Control Mechanisms in Blood-Feeding Arthropods. J Innate Immun. 2011;3(1):41-51.

741. Sojka D, Franta Z, Horn M, Caffrey CR, Mares M, Kopacek P. New insights into the machinery of blood digestion by ticks. Trends Parasitol. 2013;29(6):276-85.

742. Horn M, Nussbaumerova M, Sanda M, Kovarova Z, Srba J, Franta Z, et al. Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem Biol. 2009;16(10):1053-63.

743. Sojka D, Pytelkova J, Perner J, Horn M, Konvickova J, Schrenkova J, et al. Multienzyme degradation of host serum albumin in ticks. Ticks Tick Borne Dis. 2016;7(4):604-13.

744. Sojka D, Franta Z, Horn M, Hajdusek O, Caffrey CR, Mares M, et al. Profiling of proteolytic enzymes in the gut of the tick Ixodes ricinus reveals an evolutionarily conserved network of aspartic and cysteine peptidases. Parasit Vectors. 2008;1(1):7.

745. Franta Z, Frantová H, Konvičková J, Horn M, Sojka D, Mareš M, et al. Dynamics of digestive proteolytic system during blood feeding of the hard tick Ixodes ricinus. Parasit Vectors. 2010;3:119.

746. Huntley JF, Machell J, Nisbet AJ, van den Broek A, Chua KY, Cheong N, et al. Identification of tropomyosin, paramyosin and apolipophorin/vitellogenin as three major allergens of the sheep scab mite, Psoroptes ovis. Parasite immunology. 2004;26(8-9):335-42.

747. Perry SV. Vertebrate tropomyosin: distribution, properties and function. J Muscle Res Cell M. 2001;22(1):5-49.

748. Sereda MJ, Hartmann S, Lucius R. Helminths and allergy: the example of tropomyosin. Trends Parasitol. 2008;24(6):272-8.

749. Elfvin M, Levine RJC, Dewey MM. Paramyosin in Invertebrate Muscles .1. Identification and Localization. J Cell Biol. 1976;71(1):261-72.

750. Hartmann S, Adam R, Marti T, Kirsten C, Seidinger S, Lucius R. A 41-kDa antigen of the rodent filaria Acanthocheilonema viteae with homologies to tropomyosin induces host-protective immune responses. Parasitology research. 1997;83(4):390-3.

751. Tian MY, Tian ZC, Luo J, Xie JR, Yin H, Zeng QY, et al. Identification of the tropomyosin (HL-Tm) in Haemaphysalis longicornis. Veterinary parasitology. 2015;207(3-4):318-23.

752. Ramirez BL, Kurtis JD, Wiest PM, Arias P, Aligui F, Acosta L, et al. Paramyosin: A candidate vaccine antigen against Schistosoma japonicum. Parasite immunology. 1996;18(1):49-52.

753. Mizuno H, Hamada A, Shimada K, Honda H. Tropomyosin as a regulator of the sliding movement of actin filaments. Biosystems. 2007;90(2):449-55.

754. Asturias JA, Arilla MC, Gomez-Bayon N, Martinez A, Martinez J, Palacios R. Sequencing and high level expression in Escherichia coli of the tropomyosin allergen (Der p 10) from Dermatophagoides pteronyssinus. Bba-Gene Struct Expr. 1998;1397(1):27-30.

755. Aki T, Kodama T, Fujikawa A, Miura K, Shigeta S, Wada T, et al. Immunochemical Characterization of Recombinant and Native Tropomyosins as a New Allergen from the House-Dust Mite, Dermatophagoides-Farinae. J Allergy Clin Immun. 1995;96(1):74-83.

756. Westritschnig K, Sibanda E, Thomas W, Auer H, Aspock H, Pittner G, et al. Analysis of the sensitization profile towards allergens in central Africa. Clin Exp Allergy. 2003;33(1):22-7.

757. Zhang DM, Pan WQ, Qian L, Duke M, Shen LH, McManus DP. Investigation of recombinant Schistosoma japonicum paramyosin fragments for immunogenicity and vaccine efficacy in mice. Parasite immunology. 2006;28(3):77-84.

758. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402.

759. Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, et al. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 2005;272(20):5101-9.

760. Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 2004;32(Web Server issue):W327-31.

761. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 2011;39(Database issue):D225-9.

762. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43(Database issue):D222-6.

763. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45(D1):D200-D3.

764. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

765. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. PLoS Genet. 2008;4(12):e1000304.

766. Zhong R, Ding TB, Niu JZ, Xia WK, Liao CY, Dou W, et al. Molecular characterization of vitellogenin and its receptor genes from citrus red mite, Panonychus citri (McGregor). Int J Mol Sci. 2015;16(3):4759-73.

767. Khalil SM, Donohue KV, Thompson DM, Jeffers LA, Ananthapadmanaban U, Sonenshine DE, et al. Full-length sequence, regulation and developmental studies of a second vitellogenin gene from the American dog tick, Dermacentor variabilis. J Insect Physiol. 2011;57(3):400-8.

768. Horigane M, Shinoda T, Honda H, Taylor D. Characterization of a vitellogenin gene reveals two phase regulation of vitellogenesis by engorgement and mating in the soft tick Ornithodoros moubata (Acari: Argasidae). Insect Mol Biol. 2010;19(4):501-15.

769. Kawakami Y, Goto SG, Ito K, Numata H. Suppression of ovarian development and vitellogenin gene expression in the adult diapause of the two-spotted spider mite Tetranychus urticae. J Insect Physiol. 2009;55(1):70-7.

770. Guo G, Wang L, Wu H, Cao Y, Xiao R, Lai X, et al. Molecular characterization and expression of vitellogenin genes from the wolf spider P*ardosa pseudoannulata* (Araneae: Lycosidae). Physiological Entomology. 2018;43:295-305.

771. Zhao Y, Li D, Zhang M, Chen W, Zhang G. Food source affects the expression of vitellogenin and fecundity of a biological control agent, Neoseiulus cucumeris. Exp Appl Acarol. 2014;63(3):333-47.

772. Cabrera Cordon AR, Shirk PD, Duehl AJ, Evans JD, Teal PE. Variable induction of vitellogenin genes in the varroa mite, Varroa destructor (Anderson & Trueman), by the honeybee, Apis mellifera L, host and its environment. Insect Mol Biol. 2013;22(1):88-103.

773. Tirloni L, Seixas A, Mulenga A, Vaz IaS, Termignoni C. A family of serine protease inhibitors (serpins) in the cattle tick Rhipicephalus (Boophilus) microplus. Exp Parasitol. 2014;137:25-34.

774. Liu X, Shen G, Xu H, He L. The fenpropathrin resistant Tetranychus cinnabarinus showed increased fecundity with high content of vitellogenin and vitellogenin receptor. Pestic Biochem Physiol. 2016;134:31-8.

775. de Castro MH, de Klerk D, Pienaar R, Latif AA, Rees DJ, Mans BJ. De novo assembly and annotation of the salivary gland transcriptome of Rhipicephalus appendiculatus male and female ticks during blood feeding. Ticks Tick Borne Dis. 2016;7(4):536-48.

776. Dong X, Armstrong SD, Xia D, Makepeace BL, Darby AC, Kadowaki T. Draft genome of the honey bee ectoparasitic mite, Tropilaelaps mercedesae, is shaped by the parasitic life history. Gigascience. 2017;6(3):1-17.

777. Hughes AL. Evolutionary conservation of amino acid composition in paralogous insect vitellogenins. Gene. 2010;467(1-2):35-40.

778. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987.

779. de Jong MA, Wahlberg N, van Eijk M, Brakefield PM, Zwaan BJ. Mitochondrial DNA signature for range-wide populations of Bicyclus anynana suggests a rapid expansion from recent refugia. PLoS One. 2011;6(6):e21385.

780. Ribeiro JMC, Genta FA, Sorgine MHF, Logullo R, Mesquita RD, Paiva-Silva GO, et al. An Insight into the Transcriptome of the Digestive Tract of the Bloodsucking Bug, Rhodnius prolixus. Plos Neglect Trop D. 2014;8(1).

781. Sojka D, Franta Z, Frantová H, Bartosová P, Horn M, Váchová J, et al. Characterization of gut-associated cathepsin D hemoglobinase from tick Ixodes ricinus (IrCD1). J Biol Chem. 2012;287(25):21152-63.

782. Andreson R, Mols T, Remm M. Predicting failure rate of PCR in large genomes. Nucleic Acids Res. 2008;36(11):e66.

783. Beasley EM, Myers RM, Cox DR, Lazzeroni LC, Innis MA, Gelfand DH, et al. Statistical refinement of primer design parameters. PCR Applications: Protocols for Functional Genomics. 1999;1:55-72.
784. Varadaraj K, Skinner DM. Denaturants or Cosolvents Improve the Specificity of Pcr Amplification of a G+C-Rich DNA Using Genetically-Engineered DNA-Polymerases. Gene. 1994;140(1):1-5.
785. Haas S, Vingron M, Poustka A, Wiemann S. Primer design for large scale sequencing. Nucleic Acids Research. 1998;26(12):3006-12.
786. Cobb BD, Clarkson JM. A Simple Procedure for Optimizing the Polymerase Chain-Reaction (Pcr) Using Modified Taguchi Methods. Nucleic Acids Research. 1994;22(18):3801-5.
787. Petrov D. [Study of Dermanyssus gallinae as a carrier of Pasteurella multocida]. Vet Med Nauki. 1975;12(5):32-6.

10 SUPPLEMENTARY

10.1 MITE SAMPLING PACK CONTENTS

10.1.1 Questionnaire regarding *D. gallinae* in the United Kingdom

Questionnaire regarding Poultry red mite in the United Kingdom

The following questionnaire forms part of a PhD project conducted by Eleanor Karp-Tatham, supervised by Prof. Damer Blake, Prof. Fiona Tomley, Dr Tatiana Küster and Prof. Alasdair Nisbet, conducted at the Royal Veterinary College and Moredun Institute.

Please read prior to completing the questionnaire:

- Completed questionnaires will be treated confidentially. Anonymity will be ensured throughout and details concerning farms will not be shared with any third party.
- Please answer all questions.

BASIC FARM DETAILS:

1) Farm name:
2) Owner/Manager:
3) Address:
4) Is your production system?
   - ☐ Free range
   - ☐ Non free-range (indoor only)
   - ☐ Other (please specify) ....
5) Do you operate?
   - ☐ Organically
   - ☐ Non-organically
6) When was the farm established? ……………………

7) How many units are present on the site?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5+

8) What is the stocking density of the units present? ……………………………..

9) How old are the units?

- ☐ Up to 1 year
- ☐ 2 years old
- ☐ 3 years old
- ☐ 4 years old
- ☐ Other (please specify)

  ……………………………………………………………………………………………………….

10) What is the total number of chickens present on the farm?

    ………………………………………………….

11) How far away is the closest farm housing poultry?

    …………………………………………………………………..

12) Are you in contact with local farms? If so, are you aware if they are affected by red mite also?

- ☐ Yes
  - o Affected
  - o Not affected
- ☐ No

<u>CHICKENS</u>:

1) Breed of chicken used currently

   ……………………………………………………………………………………………

2) Current age of chickens

.... Years .... Months

3) Who is your supplier of chickens?

....

4) Do you notice a correlation between age of chicken and level of red mite infestation? (i.e. does the number of poultry red mite increase with the age of chicken)

- ☐ Yes
- ☐ No

<u>HOUSING AND FEEDING:</u>

1) Type of housing system used?
   - ☐ Enriched cage
   - ☐ Multi-tier aviary
   - ☐ Single-tier aviary
   - ☐ Battery cage
   - ☐ Other (please specify) ....
2) Does your system utilise a:
   - ☐ Manure belt
   - ☐ Manure pit
   - ☐ Deep litter
   - ☐ Other (please specify).................
3) What brand of litter do you use? ....
4) Have you ever changed you type of litter?
   - ☐ Yes

     Reason ....
   - ☐ No
5) What diet do you feed to your chickens? ....
6) Do you use any feed additives?
   - ☐ Yes (please specifiy) ....
   - ☐ No
7) If free-range, do you have wooden and/or metal structures outside?

- ☐ Yes, wooden
- ☐ Yes, metal
- ☐ Yes, both
- ☐ None

<u>Dermanyssus gallinae:</u>

1) Have you noticed red mite on your farm?
   - ☐ Yes
   - ☐ No
2) If so, when were you first aware of the mite?

   …………………………………………………………………………

3) Do you notice seasonal fluctuations in mite numbers?
   - ☐ Yes
   - ☐ No
4) Do you feel that red mite is a significant problem?
   - ☐ Yes, significant
   - ☐ Yes, but moderate
   - ☐ No
5) Do you feel like this is sufficient information available about red mite?
   - ☐ Yes
   - ☐ No

6) Are there any clear indicators of a heavy infestation? (please provide details)

7) What are the major problems you associate with red mite? (please provide details)

CONTROL OF RED MITE

1) What control measures are in place on your farm for red mite? (please tick all that apply)
   - ☐ Chemical (e.g. Acaricides)
   - ☐ Silica powder and/or desiccant dust
   - ☐ Hygiene treatments
   - ☐ Heat/cold treatment
   - ☐ Natural products
   - ☐ Feed additives
   - ☐ Other (please specify) ……………………………………
2) How often do you use each control method? (Please provide details)

3) How do you clean the empty poultry houses?

4) What specific product(s) do you use in the control of red mite? (e.g. specific brands)

5) Do you feel like your current control strategies are effective?
   - ☐ Yes
   - ☐ No
6) What have you used previously?

7) Have you considered using anything else? If so, why have you chosen not to use it yet?

- ☐ Cost
- ☐ Ease of application
- ☐ Other (please specify below)

If you have any further comments, please feel free to write on the attached blank sheet.

For further information, please contact Eleanor Karp-Tatham on ekarp@rvc.ac.uk

Thank you for your participation

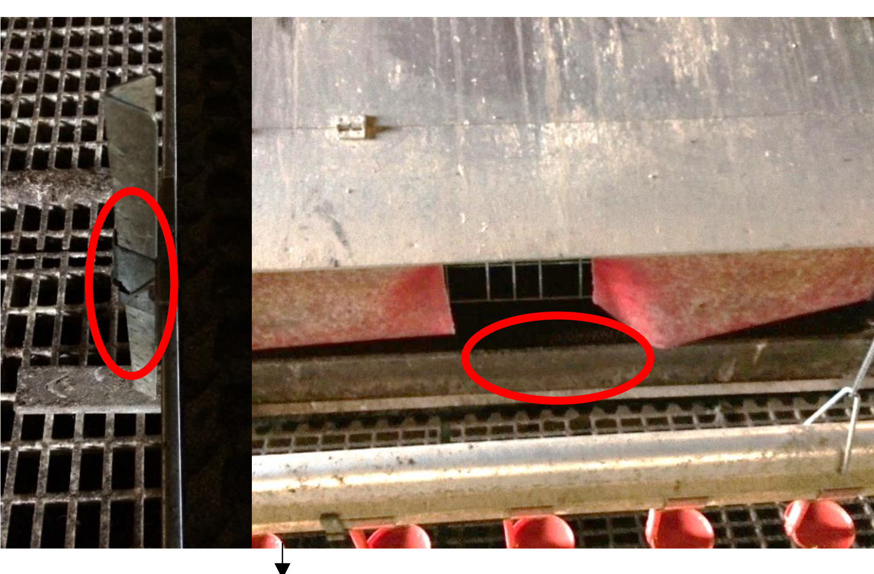
10.1.2 Instructions for setting traps for *D. gallinae*

Instructions for setting mite traps

1. Remove traps and place inside barns, in locations where mites are normally found

Common places I find mites when I go mite collecting is underneath the mats in the nest boxes or down the sides of the feeder legs (see photos). Feel free to place traps where you think best!

(Apologies for the quality of photos)



NOTE: If your farm has multiple barns it would be great if you could distribute traps in the two furthest apart! This should help me find out if there is diversity present within the mites from different barns on

2. Leave traps in position for 7-14 days

3. Remove traps and place back into plastic bag provided. If you have placed traps in separate barns, an additional bag is provided

Please place traps from one barn per bag, to avoid mixing mites.

Please write on the label; Barn, location mite trap was placed (e.g. nest box) and date the trap was

4. Place plastic bags from step 3 into larger ziplock bag provided

Please seal tightly! It is important that the bags are shut as tightly as possible to avoid the possibility of mites escaping from the envelope during transit.

5. Place mite traps and completed questionnaire into the provided envelope

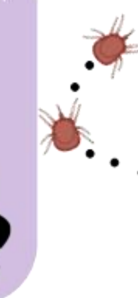
Envelopes come with address attached and postage already paid for!

10.1.3 Information sheet provided to UK farmers

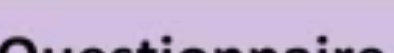
Are poultry red mites more diverse than they appear? We want to know!

Who are you helping?

Eleanor Karp-Tatham	Prof. Damer Blake	Prof. Fiona Tomley	Dr Tatiana Küster	Prof. Alasdair Nisbet



PhD Student



Supervisory Team

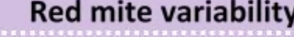
Why do we need mite samples?

Project aim: To gain a better understanding of poultry red mite variability to help develop effective control strategies

Where are we targeting?

- Across the UK
- Across Europe and worldwide samples
- Studying diversity between farms and within the same farm

Red mite variability



- Poultry red mites look almost identical to the naked eye but could be variable at a molecular level
- High variability in mites would allow fast adaptation, for example, resistance to acaricides
- Understanding variability will help us to improve future control methods
- The mites you provide will be used directly to assist this project

Questionnaire

- Given to UK farmers
- Knowledge and opinions gained directly from farmers are valuable to assist laboratory work
- Help to provide relevant information on current control methods used
- Help to highlight common risk factors to scientists
- Knowledge will be fed back to farmers to improve awareness

Vaccine development

- Successful development of a vaccine to control poultry red mite requires knowledge on variability
- Low levels of variability would allow production of a single vaccine
- High levels of variability would suggest we need to develop vaccines targeted at specific mite populations

10.2 Statistical output regarding nucleotide and haplotype diversity for COI fragments

10.2.1 Nucleotide diversity Dunnett's multiple comparison test

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value
All samples vs. UK	0.01157	0.01029 to 0.01285	Yes	****	<0.0001
All samples vs. Greece	0.02141	0.02034 to 0.02248	Yes	****	<0.0001
All samples vs. Albania	0.004360	0.001995 to 0.006725	Yes	****	<0.0001
All samples vs. Belgium	0.005690	0.003059 to 0.008321	Yes	****	<0.0001
All samples vs. Croatia	0.02241	0.01911 to 0.02571	Yes	****	<0.0001
All samples vs. Czech Republic	0.01031	0.007945 to 0.01267	Yes	****	<0.0001
All samples vs. Denmark	0.01043	0.007943 to 0.01292	Yes	****	<0.0001
All samples vs. France	0.02407	0.02105 to 0.02709	Yes	****	<0.0001
All samples vs. Italy	0.02215	0.01966 to 0.02464	Yes	****	<0.0001
All samples vs. Portugal	0.003690	0.001325 to 0.006055	Yes	****	<0.0001
All samples vs. Romania	0.02155	0.01874 to 0.02436	Yes	****	<0.0001
All samples vs. Slovenia	-0.0007000	-0.003506 to 0.002106	No	ns	>0.9999
All samples vs. Turkey	0.02501	0.02199 to 0.02803	Yes	****	<0.0001
All samples vs. Netherlands	3.000e-005	-0.002457 to 0.002517	No	ns	>0.9999
UK vs. Greece	0.009840	0.008344 to 0.01134	Yes	****	<0.0001
UK vs. Albania	-0.007210	-0.009796 to -0.004624	Yes	****	<0.0001
UK vs. Belgium	-0.005880	-0.008711 to -0.003049	Yes	****	<0.0001
UK vs. Croatia	0.01084	0.007375 to 0.01431	Yes	****	<0.0001
UK vs. Czech Republic	-0.001260	-0.003846 to 0.001326	No	ns	0.9426
UK vs. Denmark	-0.001140	-0.003838 to 0.001558	No	ns	0.9828
UK vs. France	0.01250	0.009301 to 0.01570	Yes	****	<0.0001
UK vs. Italy	0.01058	0.007882 to 0.01328	Yes	****	<0.0001
UK vs. Portugal	-0.007880	-0.01047 to -0.005294	Yes	****	<0.0001

UK vs. Romania	0.009980	0.006986 to 0.01297	Yes	****	<0.0001
UK vs. Slovenia	-0.01227	-0.01526 to -0.009276	Yes	****	<0.0001
UK vs. Turkey	0.01344	0.01024 to 0.01664	Yes	****	<0.0001
UK vs. Netherlands	-0.01154	-0.01424 to -0.008842	Yes	****	<0.0001
Greece vs. Albania	-0.01705	-0.01954 to -0.01456	Yes	****	<0.0001
Greece vs. Belgium	-0.01572	-0.01846 to -0.01298	Yes	****	<0.0001
Greece vs. Croatia	0.001000	-0.002393 to 0.004393	No	ns	0.9996
Greece vs. Czech Republic	-0.01110	-0.01359 to -0.008611	Yes	****	<0.0001
Greece vs. Denmark	-0.01098	-0.01358 to -0.008375	Yes	****	<0.0001
Greece vs. France	0.002660	-0.0004612 to 0.005781	No	ns	0.1967
Greece vs. Italy	0.0007400	-0.001865 to 0.003345	No	ns	0.9997
Greece vs. Portugal	-0.01772	-0.02021 to -0.01523	Yes	****	<0.0001
Greece vs. Romania	0.0001400	-0.002771 to 0.003051	No	ns	>0.9999
Greece vs. Slovenia	-0.02211	-0.02502 to -0.01920	Yes	****	<0.0001
Greece vs. Turkey	0.003600	0.0004788 to 0.006721	Yes	**	0.0084
Greece vs. Netherlands	-0.02138	-0.02398 to -0.01878	Yes	****	<0.0001
Albania vs. Belgium	0.001330	-0.002130 to 0.004790	No	ns	0.9930
Albania vs. Croatia	0.01805	0.01405 to 0.02205	Yes	****	<0.0001
Albania vs. Czech Republic	0.005950	0.002688 to 0.009212	Yes	****	<0.0001
Albania vs. Denmark	0.006070	0.002718 to 0.009422	Yes	****	<0.0001
Albania vs. France	0.01971	0.01594 to 0.02348	Yes	****	<0.0001
Albania vs. Italy	0.01779	0.01444 to 0.02114	Yes	****	<0.0001
Albania vs. Portugal	-0.0006700	-0.003932 to 0.002592	No	ns	>0.9999
Albania vs. Romania	0.01719	0.01360 to 0.02078	Yes	****	<0.0001
Albania vs. Slovenia	-0.005060	-0.008655 to -0.001465	Yes	***	0.0002
Albania vs. Turkey	0.02065	0.01688 to 0.02442	Yes	****	<0.0001
Albania vs. Netherlands	-0.004330	-0.007682 to -0.0009782	Yes	**	0.0013
Belgium vs. Croatia	0.01672	0.01256 to 0.02088	Yes	****	<0.0001
Belgium vs. Czech Republic	0.004620	0.001160 to 0.008080	Yes	***	0.0007
Belgium vs. Denmark	0.004740	0.001195 to 0.008285	Yes	***	0.0006

Belgium vs. France	0.01838	0.01444 to 0.02232	Yes	****	<0.0001
Belgium vs. Italy	0.01646	0.01292 to 0.02000	Yes	****	<0.0001
Belgium vs. Portugal	-0.002000	-0.005460 to 0.001460	No	ns	0.8128
Belgium vs. Romania	0.01586	0.01208 to 0.01964	Yes	****	<0.0001
Belgium vs. Slovenia	-0.006390	-0.01017 to -0.002615	Yes	****	<0.0001
Belgium vs. Turkey	0.01932	0.01538 to 0.02326	Yes	****	<0.0001
Belgium vs. Netherlands	-0.005660	-0.009205 to -0.002115	Yes	****	<0.0001
Croatia vs. Czech Republic	-0.01210	-0.01610 to -0.008104	Yes	****	<0.0001
Croatia vs. Denmark	-0.01198	-0.01605 to -0.007911	Yes	****	<0.0001
Croatia vs. France	0.001660	-0.002757 to 0.006077	No	ns	0.9944
Croatia vs. Italy	-0.0002600	-0.004329 to 0.003809	No	ns	>0.9999
Croatia vs. Portugal	-0.01872	-0.02272 to -0.01472	Yes	****	<0.0001
Croatia vs. Romania	-0.0008600	-0.005131 to 0.003411	No	ns	>0.9999
Croatia vs. Slovenia	-0.02311	-0.02738 to -0.01884	Yes	****	<0.0001
Croatia vs. Turkey	0.002600	-0.001817 to 0.007017	No	ns	0.7917
Croatia vs. Netherlands	-0.02238	-0.02645 to -0.01831	Yes	****	<0.0001
Czech Republic vs. Denmark	0.0001200	-0.003232 to 0.003472	No	ns	>0.9999
Czech Republic vs. France	0.01376	0.009993 to 0.01753	Yes	****	<0.0001
Czech Republic vs. Italy	0.01184	0.008488 to 0.01519	Yes	****	<0.0001
Czech Republic vs. Portugal	-0.006620	-0.009882 to -0.003358	Yes	****	<0.0001
Czech Republic vs. Romania	0.01124	0.007645 to 0.01483	Yes	****	<0.0001
Czech Republic vs. Slovenia	-0.01101	-0.01460 to -0.007415	Yes	****	<0.0001
Czech Republic vs. Turkey	0.01470	0.01093 to 0.01847	Yes	****	<0.0001
Czech Republic vs. Netherlands	-0.01028	-0.01363 to -0.006928	Yes	****	<0.0001
Denmark vs. France	0.01364	0.009795 to 0.01748	Yes	****	<0.0001
Denmark vs. Italy	0.01172	0.008281 to 0.01516	Yes	****	<0.0001
Denmark vs. Portugal	-0.006740	-0.01009 to -0.003388	Yes	****	<0.0001
Denmark vs. Romania	0.01112	0.007444 to 0.01480	Yes	****	<0.0001
Denmark vs. Slovenia	-0.01113	-0.01481 to -0.007454	Yes	****	<0.0001
Denmark vs. Turkey	0.01458	0.01074 to 0.01842	Yes	****	<0.0001

Denmark vs. Netherlands	-0.01040	-0.01384 to -0.006961	Yes	****	<0.0001
France vs. Italy	-0.001920	-0.005765 to 0.001925	No	ns	0.9306
France vs. Portugal	-0.02038	-0.02415 to -0.01661	Yes	****	<0.0001
France vs. Romania	-0.002520	-0.006579 to 0.001539	No	ns	0.7209
France vs. Slovenia	-0.02477	-0.02883 to -0.02071	Yes	****	<0.0001
France vs. Turkey	0.0009400	-0.003272 to 0.005152	No	ns	>0.9999
France vs. Netherlands	-0.02404	-0.02788 to -0.02020	Yes	****	<0.0001
Italy vs. Portugal	-0.01846	-0.02181 to -0.01511	Yes	****	<0.0001
Italy vs. Romania	-0.0006000	-0.004276 to 0.003076	No	ns	>0.9999
Italy vs. Slovenia	-0.02285	-0.02653 to -0.01917	Yes	****	<0.0001
Italy vs. Turkey	0.002860	-0.0009848 to 0.006705	No	ns	0.4148
Italy vs. Netherlands	-0.02212	-0.02556 to -0.01868	Yes	****	<0.0001
Portugal vs. Romania	0.01786	0.01427 to 0.02145	Yes	****	<0.0001
Portugal vs. Slovenia	-0.004390	-0.007985 to -0.0007950	Yes	**	0.0034
Portugal vs. Turkey	0.02132	0.01755 to 0.02509	Yes	****	<0.0001
Portugal vs. Netherlands	-0.003660	-0.007012 to -0.0003082	Yes	*	0.0178
Romania vs. Slovenia	-0.02225	-0.02615 to -0.01835	Yes	****	<0.0001
Romania vs. Turkey	0.003460	-0.0005985 to 0.007519	No	ns	0.1962
Romania vs. Netherlands	-0.02152	-0.02520 to -0.01784	Yes	****	<0.0001
Slovenia vs. Turkey	0.02571	0.02165 to 0.02977	Yes	****	<0.0001
Slovenia vs. Netherlands	0.0007300	-0.002946 to 0.004406	No	ns	>0.9999
Turkey vs. Netherlands	-0.02498	-0.02882 to -0.02114	Yes	****	<0.0001

Table 56: *Dunnett's comparison test for nucleotide diversity on a fragment of the COI gene for* D. gallinae *isolates*

10.2.2 Haplotype diversity Tukeys multiple comparison test

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value
All samples vs. UK	0.01600	-0.02137 to 0.05337	No	ns	0.9807
All samples vs. Greece	0.3960	0.3648 to 0.4272	Yes	****	<0.0001
All samples vs. Albania	0.02800	-0.04109 to 0.09709	No	ns	0.9883
All samples vs. Belgium	-0.04700	-0.1239 to 0.02987	No	ns	0.7426
All samples vs. Croatia	0.01700	-0.07952 to 0.1135	No	ns	>0.9999
All samples vs. Czech Republic	-0.01600	-0.08509 to 0.05309	No	ns	>0.9999
All samples vs. Denmark	0.05600	-0.01665 to 0.1287	No	ns	0.3523
All samples vs. France	0.1840	0.09567 to 0.2723	Yes	****	<0.0001
All samples vs. Italy	0.1950	0.1223 to 0.2677	Yes	****	<0.0001
All samples vs. Portugal	0.02800	-0.04109 to 0.09709	No	ns	0.9883
All samples vs. Romania	0.6310	0.5490 to 0.7130	Yes	****	<0.0001
All samples vs. Slovenia	0.06000	-0.02198 to 0.1420	No	ns	0.4439
All samples vs. Turkey	0.5840	0.4957 to 0.6723	Yes	****	<0.0001
All samples vs. Netherlands	-0.02700	-0.09965 to 0.04565	No	ns	0.9950
UK vs. Greece	0.3800	0.3363 to 0.4237	Yes	****	<0.0001
UK vs. Albania	0.01200	-0.06354 to 0.08754	No	ns	>0.9999
UK vs. Belgium	-0.06300	-0.1457 to 0.01972	No	ns	0.3731
UK vs. Croatia	0.001000	-0.1002 to 0.1022	No	ns	>0.9999
UK vs. Czech Republic	-0.03200	-0.1075 to 0.04354	No	ns	0.9824
UK vs. Denmark	0.04000	-0.03881 to 0.1188	No	ns	0.9217
UK vs. France	0.1680	0.07454 to 0.2615	Yes	****	<0.0001
UK vs. Italy	0.1790	0.1002 to 0.2578	Yes	****	<0.0001
UK vs. Portugal	0.01200	-0.06354 to 0.08754	No	ns	>0.9999
UK vs. Romania	0.6150	0.5275 to 0.7025	Yes	****	<0.0001
UK vs. Slovenia	0.04400	-0.04348 to 0.1315	No	ns	0.9268
UK vs. Turkey	0.5680	0.4745 to 0.6615	Yes	****	<0.0001

UK vs. Netherlands	-0.04300	-0.1218 to 0.03581	No	ns	0.8699
Greece vs. Albania	-0.3680	-0.4407 to -0.2953	Yes	****	<0.0001
Greece vs. Belgium	-0.4430	-0.5231 to -0.3629	Yes	****	<0.0001
Greece vs. Croatia	-0.3790	-0.4781 to -0.2799	Yes	****	<0.0001
Greece vs. Czech Republic	-0.4120	-0.4847 to -0.3393	Yes	****	<0.0001
Greece vs. Denmark	-0.3400	-0.4161 to -0.2639	Yes	****	<0.0001
Greece vs. France	-0.2120	-0.3032 to -0.1208	Yes	****	<0.0001
Greece vs. Italy	-0.2010	-0.2771 to -0.1249	Yes	****	<0.0001
Greece vs. Portugal	-0.3680	-0.4407 to -0.2953	Yes	****	<0.0001
Greece vs. Romania	0.2350	0.1500 to 0.3200	Yes	****	<0.0001
Greece vs. Slovenia	-0.3360	-0.4210 to -0.2510	Yes	****	<0.0001
Greece vs. Turkey	0.1880	0.09682 to 0.2792	Yes	****	<0.0001
Greece vs. Netherlands	-0.4230	-0.4991 to -0.3469	Yes	****	<0.0001
Albania vs. Belgium	-0.07500	-0.1761 to 0.02609	No	ns	0.4195
Albania vs. Croatia	-0.01100	-0.1277 to 0.1057	No	ns	>0.9999
Albania vs. Czech Republic	-0.04400	-0.1393 to 0.05131	No	ns	0.9629
Albania vs. Denmark	0.02800	-0.06992 to 0.1259	No	ns	0.9997
Albania vs. France	0.1560	0.04595 to 0.2661	Yes	***	0.0002
Albania vs. Italy	0.1670	0.06908 to 0.2649	Yes	****	<0.0001
Albania vs. Portugal	0.000	-0.09531 to 0.09531	No	ns	>0.9999
Albania vs. Romania	0.6030	0.4980 to 0.7080	Yes	****	<0.0001
Albania vs. Slovenia	0.03200	-0.07303 to 0.1370	No	ns	0.9994
Albania vs. Turkey	0.5560	0.4459 to 0.6661	Yes	****	<0.0001
Albania vs. Netherlands	-0.05500	-0.1529 to 0.04292	No	ns	0.8430
Belgium vs. Croatia	0.06400	-0.05750 to 0.1855	No	ns	0.8977
Belgium vs. Czech Republic	0.03100	-0.07009 to 0.1321	No	ns	0.9994
Belgium vs. Denmark	0.1030	-0.0005578 to 0.2066	No	ns	0.0529
Belgium vs. France	0.2310	0.1159 to 0.3461	Yes	****	<0.0001
Belgium vs. Italy	0.2420	0.1384 to 0.3456	Yes	****	<0.0001
Belgium vs. Portugal	0.07500	-0.02609 to 0.1761	No	ns	0.4195

Belgium vs. Romania	0.6780	0.5677 to 0.7883	Yes	****	<0.0001
Belgium vs. Slovenia	0.1070	-0.003300 to 0.2173	No	ns	0.0680
Belgium vs. Turkey	0.6310	0.5159 to 0.7461	Yes	****	<0.0001
Belgium vs. Netherlands	0.02000	-0.08356 to 0.1236	No	ns	>0.9999
Croatia vs. Czech Republic	-0.03300	-0.1497 to 0.08373	No	ns	0.9997
Croatia vs. Denmark	0.03900	-0.07987 to 0.1579	No	ns	0.9986
Croatia vs. France	0.1670	0.03795 to 0.2961	Yes	**	0.0012
Croatia vs. Italy	0.1780	0.05913 to 0.2969	Yes	****	<0.0001
Croatia vs. Portugal	0.01100	-0.1057 to 0.1277	No	ns	>0.9999
Croatia vs. Romania	0.6140	0.4892 to 0.7388	Yes	****	<0.0001
Croatia vs. Slovenia	0.04300	-0.08179 to 0.1678	No	ns	0.9977
Croatia vs. Turkey	0.5670	0.4379 to 0.6961	Yes	****	<0.0001
Croatia vs. Netherlands	-0.04400	-0.1629 to 0.07487	No	ns	0.9952
Czech Republic vs. Denmark	0.07200	-0.02592 to 0.1699	No	ns	0.4356
Czech Republic vs. France	0.2000	0.08995 to 0.3101	Yes	****	<0.0001
Czech Republic vs. Italy	0.2110	0.1131 to 0.3089	Yes	****	<0.0001
Czech Republic vs. Portugal	0.04400	-0.05131 to 0.1393	No	ns	0.9629
Czech Republic vs. Romania	0.6470	0.5420 to 0.7520	Yes	****	<0.0001
Czech Republic vs. Slovenia	0.07600	-0.02903 to 0.1810	No	ns	0.4643
Czech Republic vs. Turkey	0.6000	0.4899 to 0.7101	Yes	****	<0.0001
Czech Republic vs. Netherlands	-0.01100	-0.1089 to 0.08692	No	ns	>0.9999
Denmark vs. France	0.1280	0.01568 to 0.2403	Yes	**	0.0100
Denmark vs. Italy	0.1390	0.03853 to 0.2395	Yes	***	0.0003
Denmark vs. Portugal	-0.02800	-0.1259 to 0.06992	No	ns	0.9997
Denmark vs. Romania	0.5750	0.4676 to 0.6824	Yes	****	<0.0001
Denmark vs. Slovenia	0.004000	-0.1034 to 0.1114	No	ns	>0.9999
Denmark vs. Turkey	0.5280	0.4157 to 0.6403	Yes	****	<0.0001
Denmark vs. Netherlands	-0.08300	-0.1835 to 0.01747	No	ns	0.2405
France vs. Italy	0.01100	-0.1013 to 0.1233	No	ns	>0.9999
France vs. Portugal	-0.1560	-0.2661 to -0.04595	Yes	***	0.0002

France vs. Romania	0.4470	0.3284 to 0.5656	Yes	****	<0.0001
France vs. Slovenia	-0.1240	-0.2426 to -0.005431	Yes	*	0.0304
France vs. Turkey	0.4000	0.2770 to 0.5230	Yes	****	<0.0001
France vs. Netherlands	-0.2110	-0.3233 to -0.09868	Yes	****	<0.0001
Italy vs. Portugal	-0.1670	-0.2649 to -0.06908	Yes	****	<0.0001
Italy vs. Romania	0.4360	0.3286 to 0.5434	Yes	****	<0.0001
Italy vs. Slovenia	-0.1350	-0.2424 to -0.02760	Yes	**	0.0021
Italy vs. Turkey	0.3890	0.2767 to 0.5013	Yes	****	<0.0001
Italy vs. Netherlands	-0.2220	-0.3225 to -0.1215	Yes	****	<0.0001
Portugal vs. Romania	0.6030	0.4980 to 0.7080	Yes	****	<0.0001
Portugal vs. Slovenia	0.03200	-0.07303 to 0.1370	No	ns	0.9994
Portugal vs. Turkey	0.5560	0.4459 to 0.6661	Yes	****	<0.0001
Portugal vs. Netherlands	-0.05500	-0.1529 to 0.04292	No	ns	0.8430
Romania vs. Slovenia	-0.5710	-0.6849 to -0.4571	Yes	****	<0.0001
Romania vs. Turkey	-0.04700	-0.1656 to 0.07157	No	ns	0.9906
Romania vs. Netherlands	-0.6580	-0.7654 to -0.5506	Yes	****	<0.0001
Slovenia vs. Turkey	0.5240	0.4054 to 0.6426	Yes	****	<0.0001
Slovenia vs. Netherlands	-0.08700	-0.1944 to 0.02040	No	ns	0.2704
Turkey vs. Netherlands	-0.6110	-0.7233 to -0.4987	Yes	****	<0.0001

Table 57: *Tukey's comparison test for haplotype diversity on a fragment of the COI gene for D. gallinae isolates*

10.2.3 Haplotype and nucleotide diversity comparison of Greece and United Kingdom

10.2.3.1 Nucleotide diversity comparison

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	
All samples vs. UK	0.01157	0.01101 to 0.01213	Yes	****	<0.0001	A-B
All samples vs. Greece	0.02141	0.02094 to 0.02188	Yes	****	<0.0001	A-C
UK vs. Greece	0.009840	0.009182 to 0.01050	Yes	****	<0.0001	B-C

Table 58: *Tukey's multiple comparison test comparing nucleotide diversity for UK, Greece and full dataset for fragment of the COI gene for* D. gallinae

10.2.3.2 Haplotype diversity comparison

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	
All samples vs. UK	0.01600	0.005966 to 0.02603	Yes	***	0.0006	A-B
All samples vs. Greece	0.3960	0.3876 to 0.4044	Yes	****	<0.0001	A-C
UK vs. Greece	0.3800	0.3683 to 0.3917	Yes	****	<0.0001	B-C

Table 59: *Tukey's multiple comparison test comparing haplotype diversity for UK, Greece and full dataset for fragment of the COI gene for* D. gallinae

10.2.4 Nucleotide and haplotype diversity comparison between UK and European countries

10.2.4.1 Nucleotide diversity comparison: UK to individual countries

Sidak's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	B-?	
UK vs. All samples	-0.01157	-0.01267 to -0.01047	Yes	****	<0.0001	A	All samples
UK vs. Greece	0.009840	0.008559 to 0.01112	Yes	****	<0.0001	C	Greece
UK vs. Albania	-0.007210	-0.009425 to -0.004995	Yes	****	<0.0001	D	Albania
UK vs. Belgium	-0.005880	-0.008305 to -0.003455	Yes	****	<0.0001	E	Belgium
UK vs. Croatia	0.01084	0.007872 to 0.01381	Yes	****	<0.0001	F	Croatia
UK vs. Czech Republic	-0.001260	-0.003475 to 0.0009550	No	ns	0.7604	G	Czech Republic
UK vs. Denmark	-0.001140	-0.003451 to 0.001171	No	ns	0.8971	H	Denmark
UK vs. France	0.01250	0.009760 to 0.01524	Yes	****	<0.0001	I	France
UK vs. Italy	0.01058	0.008269 to 0.01289	Yes	****	<0.0001	J	Italy
UK vs. Portugal	-0.007880	-0.01010 to -0.005665	Yes	****	<0.0001	K	Portugal
UK vs. Romania	0.009980	0.007415 to 0.01255	Yes	****	<0.0001	L	Romania
UK vs. Slovenia	-0.01227	-0.01484 to -0.009705	Yes	****	<0.0001	M	Slovenia
UK vs. Turkey	0.01344	0.01070 to 0.01618	Yes	****	<0.0001	N	Turkey
UK vs. Netherlands	-0.01154	-0.01385 to -0.009229	Yes	****	<0.0001	O	Netherlands

Table 60*: Sidak's multiple comparison test for comparison of nucleotide diversity of individual countries to the UK for fragment of the COI gene for* D. gallinae

10.2.4.2 Haplotype diversity comparison

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	B-?	
UK vs. All samples	-0.01600	-0.04775 to 0.01575	No	ns	0.8420	A	All samples
UK vs. Greece	0.3800	0.3429 to 0.4171	Yes	****	<0.0001	C	Greece
UK vs. Albania	0.01200	-0.05219 to 0.07619	No	ns	0.9994	D	Albania
UK vs. Belgium	-0.06300	-0.1333 to 0.007288	No	ns	0.1158	E	Belgium
UK vs. Croatia	0.001000	-0.08502 to 0.08702	No	ns	>0.9999	F	Croatia
UK vs. Czech Republic	-0.03200	-0.09619 to 0.03219	No	ns	0.8512	G	Czech Republic
UK vs. Denmark	0.04000	-0.02697 to 0.1070	No	ns	0.6457	H	Denmark
UK vs. France	0.1680	0.08858 to 0.2474	Yes	****	<0.0001	I	France
UK vs. Italy	0.1790	0.1120 to 0.2460	Yes	****	<0.0001	J	Italy
UK vs. Portugal	0.01200	-0.05219 to 0.07619	No	ns	0.9994	K	Portugal
UK vs. Romania	0.6150	0.5407 to 0.6893	Yes	****	<0.0001	L	Romania
UK vs. Slovenia	0.04400	-0.03034 to 0.1183	No	ns	0.6582	M	Slovenia
UK vs. Turkey	0.5680	0.4886 to 0.6474	Yes	****	<0.0001	N	Turkey
UK vs. Netherlands	-0.04300	-0.1100 to 0.02397	No	ns	0.5405	O	Netherlands

***Table 61**: Dunnett's multiple comparison test for comparison of haplotype diversity between the UK and individual countries for fragment of the COI gene for* D. gallinae

10.2.5 Nucleotide and haplotype diversity comparison by geographical clustering

10.2.5.1 Nucleotide diversity comparison by geographical clustering

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant ?	Summary	Adjusted P Value	A-?	
All samples vs. Greece, Albania and Turkey	0.01378	0.01312 to 0.01444	Yes	****	<0.0001	B	Greece, Albania and Turkey
All samples vs. Portugal, France and Italy	0.01263	0.01158 to 0.01368	Yes	****	<0.0001	C	Portugal, France and Italy
All samples vs. Belgium and the Netherlands	0.001210	-3.545e-005 to 0.002455	No	ns	0.0597	D	Belgium and the Netherlands

Table 62: *Dunnett's multiple comparison test for comparison of nucleotide diversity by geographical clustering for fragment of the COI gene for* D. gallinae

10.2.5.2 Haplotype diversity comparison by geographical clustering

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant ?	Summary	Adjusted P Value	A-?	
All samples vs. Greece, Albania and Turkey	0.2220	0.2144 to 0.2296	Yes	****	<0.0001	B	Greece, Albania and Turkey
All samples vs. Portugal, France and Italy	0.04000	0.02801 to 0.05199	Yes	****	<0.0001	C	Portugal, France and Italy
All samples vs. Belgium and the Netherlands	-0.04600	-0.06028 to -0.03172	Yes	****	<0.0001	D	Belgium and the Netherlands

Table 63*: Dunnett's multiple comparison for comparison of haplotype diversity by geographical clustering for fragment of the COI gene for* D. gallinae

10.2.6 Nucleotide and haplotype diversity comparison of different UK production systems

10.2.6.1 Nucleotide diversity comparison of UK production systems

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	
UK vs. Free-range	0.0008000	0.0001560 to 0.001444	Yes	*	0.0110	A-B
UK vs. Intensive	-0.003100	-0.004031 to -0.002169	Yes	****	<0.0001	A-C
Free-range vs. Intensive	-0.003900	-0.004863 to -0.002937	Yes	****	<0.0001	B-C

Table 64: *Tukey's multiple comparison test for comparison of nucleotide diversity based on production system for fragment of the COI gene for* D. gallinae

10.2.6.2 Haplotype diversity comparison of UK production systems

Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	
UK vs. Free-range	0.007000	-0.01904 to 0.03304	No	ns	0.7970	A-B
UK vs. Intensive	0.05700	0.01936 to 0.09464	Yes	**	0.0015	A-C
Free-range vs. Intensive	0.05000	0.01106 to 0.08894	Yes	**	0.0083	B-C

Table 65: *Tukey's multiple comparison test for comparison of haplotype diversity by production system for fragment of the COI gene for* D. gallinae

10.2.7 Neutrality test: Sliding window comparison for the full dataset

		Fu and Li's D and F test				Tajima's D test		Nucleotide diversity	
Window	Midpoint	D*	Sign	F*	Sign	D	Sign	Pi	
1-102	52	-4.0532	**	-3.4375	**	-1.2239		0.03548	3.548
28-127	77	-4.6114	**	-3.9129	**	-1.3284		0.02724	2.724
53-152	102	-4.3316	**	-3.6133	**	-1.0839		0.02827	2.827
78-177	127	-5.5856	**	-4.5624	**	-1.2007		0.02634	2.634
103-202	152	-5.4125	**	-4.4237	**	-1.2802		0.03223	3.223
128-227	177	-6.1265	**	-4.9989	**	-1.4325		0.02915	2.915
153-252	202	-6.8405	**	-5.7308	**	-1.8946	*	0.01981	1.981
178-277	227	-6.779	**	-5.7717	**	-1.9732	*	0.01492	1.492
203-302	252	-6.6944	**	-5.8077	**	-1.9445	*	0.00873	0.873
228-327	277	-3.9908	**	-3.614	**	-1.4515		0.0129	1.29
253-352	302	-3.0738	*	-2.7324	*	-0.9946		0.01681	1.681
278-377	327	-1.5819		-1.5859		-0.9094		0.02015	2.015
303-402	352	-0.913		-0.8382		-0.3632		0.02803	2.803
328-427	377	-0.8671		-0.6735		-0.0638		0.02499	2.499
353-452	402	-1.476		-1.3395		-0.531		0.0203	2.03
378-477	427	-1.6484		-1.1802		0.1115		0.02498	2.498
403-502	452	-1.1737		-0.9794		-0.2205		0.02024	2.024
428-527	477	-2.8635	*	-2.477	*	-0.7915		0.02024	2.024
453-552	502	-3.6586	**	-3.037	**	-0.8731		0.0305	3.05
478-564	521	-4.465	**	-3.8081	**	-1.2436		0.02586	2.586

Table 66*: Sliding window output for Fu and Li's D and F test and Tajima's D for the full dataset. Significance for Fu and Li's D and F: #, P<0.10; *, P<0.05; **, P<0.02 and for Tajima's D: # P<0.10; * P<0.05; ** P<0.01; *** P<0.001*

10.2.8 Neutrality test: Sliding window comparison for Greece

		Fu and L's D and F test				Tajima's D test		Nucleotide diversity	
Window	Midpoint	D*	Sign	F*	Sign	Tajima's D	Sign	Pi	x100
1-100	50	0.5268		0.9663		1.5788		0.00478	0.478
26-125	75	0.5268		0.9663		1.5788		0.00478	0.478
51-150	100	0		0		0		0	0
76-175	125	0		0		0		0	0
101-200	150	0		0		0		0	0
126-225	175	0		0		0		0	0
151-250	200	0		0		0		0	0
176-275	225	0		0		0		0	0
201-300	250	0		0		0		0	0
226-325	275	0.5268		0.9287		1.4677		0.00459	0.459
251-350	300	0.5268		0.9287		1.4677		0.00459	0.459
276-375	325	0.5268		0.9287		1.4677		0.00459	0.459
301-400	350	0.5268		0.9287		1.4677		0.00459	0.459
326-425	375	0		0		0		0	0
351-450	400	0		0		0		0	0
376-475	425	0.7278		1.3136		2.0684	*	0.00946	0.946
401-500	450	0.7278		1.3136		2.0684	*	0.00946	0.946
426-525	475	0.7278		1.3136		2.0684	*	0.00946	0.946
451-550	500	0.8717		1.5768	#	2.4387	*	0.01432	1.432
476-565	520	0.5268		0.9818		1.6246		0.00539	0.539

Table 67: *Sliding window output for Fu and Li's D and F test and Tajima's D for Greece. Significance for Fu and Li's D and F: #, P<0.10; *, P<0.05; **, P<0.02 and for Tajima's D: # P<0.10; * P<0.05; ** P<0.01; *** P<0.001*