



McIntyre, Jennifer Ruth (2020) *Genetic markers of anthelmintic resistance in gastrointestinal parasites of ruminants*. PhD thesis.

<https://theses.gla.ac.uk/79021/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

# Genetic Markers of Anthelmintic Resistance in Gastrointestinal Parasites of Ruminants

Jennifer Ruth McIntyre  
BVM&S BVetSci(Hons) MRCVS

Submitted in fulfilment of the requirements for the degree  
of Doctor of Philosophy



Institute of Biodiversity, Animal Health and Comparative Medicine  
College of Medical, Veterinary and Life Sciences

February 2020

© Jennifer Ruth McIntyre, 2020

## Abstract

Parasitic gastroenteritis is the primary production limiting disease of sheep in the UK and is a considerable welfare concern. A global problem, it is caused by nematode parasites and mixed species infections can be common. In the UK, the primary pathogen in growing lambs is *Teladorsagia circumcincta*, an abomasal parasite of small ruminants, causing severe pathology and reduced weight gain. *T. circumcincta* is expertly adapted to both the host and the farming year and control is extremely difficult. The majority of UK farmers will use anthelmintics to manage parasitic gastroenteritis. Nevertheless, anthelmintic resistance is increasing, reducing control options. Many farmers will now dose sheep with a macrocyclic lactone (e.g. ivermectin) to treat *T. circumcincta* as this species has developed resistance to multiple anthelmintic classes. Unfortunately, over fifty percent of farms in recent UK studies had detectable ivermectin resistance.

There is a pressing need to conserve anthelmintics for future use. However, the mechanism of ivermectin resistance is unknown, and the lack of a sensitive test for ivermectin resistance limits research into resistance spread and development. Many excellent studies have investigated ivermectin resistance in nematode parasites, however mutations responsible for ivermectin resistance remain elusive. The purpose of this PhD was to perform a genome wide association study to identify genomic regions under ivermectin selection within UK *T. circumcincta* field populations. L3 progeny were sequenced pre- and post-ivermectin treatment using next generation sequencing techniques (ddRAD-Seq and Pool-Seq) and population genetics analyses were performed.

Multiple loci were genetically differentiated between pre- and post-ivermectin populations. However, the reference genomes used were highly fragmented and the number of loci under selection cannot be concluded. Genes identified included those with neuronal functions, metabolic and regulatory genes. Many genes had associations with pharyngeal structures and chemosensory behaviour. Nevertheless, multiple copies of genes expected to be single copy were detected in both reference genomes and these may have affected read alignment and results. The work performed here provides an important basis for future studies, and has generated high quality next generation sequenced resources from two UK field populations of *T. circumcincta*.

## Table of Contents

List of Tables.....	8
List of Figures.....	10
List of Accompanying Material.....	12
Acknowledgement.....	13
Author’s Declaration.....	14
Abbreviations.....	15
1 Introduction.....	19
1.1 Sheep Farming Systems.....	19
1.2 Parasitic Gastroenteritis.....	21
1.3 Teladorsagiosis.....	22
1.4 Management of Parasitic Gastroenteritis.....	24
1.5 Anthelmintic Resistance.....	28
1.6 <i>In vitro</i> tests of anthelmintic sensitivity.....	29
1.7 Prevalence of Anthelmintic Resistance.....	33
1.8 Ivermectin Resistance.....	35
1.8.1 IVM resistance is inherited as a fully or partially dominant trait....	35
1.8.2 IVM affects motility, fecundity and reproduction.....	36
1.8.3 IVM interacts with ligand gated ion channels, but the cause of resistance is unknown.....	37
1.9 Benzimidazole Resistance.....	44
1.10 Levamisole Resistance.....	45
1.11 Monepantel Resistance.....	48
1.12 What factors affect the genetic footprint of resistance?.....	50
1.13 Why might anthelmintic resistance have arisen so rapidly?.....	51
1.14 What are the concerns with anthelmintic resistance?.....	52
1.15 Is there hope in the face of anthelmintic resistance?.....	53
1.16 Investigating the genetic basis of drug resistance using next generation sequencing.....	55
2 Materials and Methods.....	61
2.1 Parasite material.....	61
2.1.1 Laboratory isolates.....	61
<i>MTci2 (Moredun T. circumcincta 2)</i> .....	61
<i>MTci5 (Moredun T. circumcincta 5)</i> .....	61
2.1.2 Farm 1.....	61
2.1.3 Farm 2.....	61
2.1.4 Farm 3.....	62
2.2 Sample Collection and Processing.....	62
2.2.1 Collection of faecal samples.....	62

2.2.2	Faecal egg counts.....	62
2.2.3	Coprocultures and Baermannisation of L3 .....	63
2.2.4	Morphological examination of L3.....	64
2.2.5	Faecal egg count reduction test.....	64
2.3	Bioassays.....	64
2.3.1	Egg hatch test.....	65
2.3.2	Larval development test.....	66
2.4	Molecular Biology.....	67
2.4.1	Lysate protocol .....	67
2.4.2	Speciation of strongyles by PCR .....	67
2.4.3	Microsatellite PCR and Analysis .....	69
2.4.4	Pyrosequencing .....	72
2.4.5	<i>hprt-1</i> allele specific PCR .....	74
2.5	Genomic DNA Library Preparation for Next Generation Sequencing: ddRAD-Seq and Pool-Seq.....	76
2.5.1	Whole Genome Amplification of DNA for ddRAD-Seq.....	76
2.5.2	Double digest of WGA DNA .....	77
2.5.3	Adaptor ligation.....	77
2.5.4	Size selection, Illumina Library Prep and DNA Sequencing.....	78
2.5.5	Preparation of Genomic DNA Libraries for Pool-Seq.....	78
2.6	Bioinformatics: Reference genomes .....	79
2.7	Bioinformatics: ddRAD-Seq.....	80
2.7.1	Cleaning and demultiplexing of ddRAD-Seq reads .....	80
2.7.2	Alignment of ddRAD-Seq reads to <i>Tci2</i> genome.....	80
2.7.3	Formation and filtering of RADloci.....	81
2.7.4	Population structure and Kinship analysis .....	81
2.7.5	Hardy Weinberg analysis.....	82
2.7.6	Linkage disequilibrium analysis.....	82
2.7.7	SNP density across the genome.....	83
2.7.8	ddRAD-Seq genetic differentiation analysis .....	83
2.8	Bioinformatics: Preliminary annotation of <i>Tci2</i> genome.....	84
2.9	Bioinformatics: Pool-Seq .....	85
2.9.1	Alignment of Pool-Seq reads to the reference genome.....	85
2.9.2	Coverage .....	86
2.9.3	Within population diversity and Tajima's D .....	86
2.9.4	Pool-Seq genetic differentiation analysis .....	87
2.9.5	Identifying genes within genetically differentiated regions.....	89
2.9.6	Obtaining gene ontology information for identified genes .....	90
2.9.7	Identifying SNPs of high or moderate impact.....	91

2.9.8	Comparison with other GWAS .....	91
2.9.9	Copy number variation.....	91
3	Don't put all your eggs in one basket: The role of diversity in PGE and anthelmintic resistance on a commercial farm .....	95
3.1	Abstract .....	95
3.2	Introduction .....	96
3.3	Results .....	102
3.3.1	Management, parasite abundance and diversity over a year on a farm	102
3.3.1.1	Study farm and flock management.....	102
3.3.1.2	Study farm anthelmintic treatments .....	103
3.3.1.3	Faecal egg counts over the year .....	105
3.3.1.4	Strongyle species present on the farm .....	111
3.3.2	Determining anthelmintic efficacy using available tests .....	114
3.3.2.1	Faecal egg count reduction test.....	114
3.3.2.2	Bioassays .....	117
3.3.2.3	Pyrosequencing.....	127
3.3.3	Assessing population change over time .....	130
3.4	Discussion .....	134
3.4.1	Parasite abundance and diversity on a farm today .....	134
3.4.2	Was the anthelmintic use appropriate? .....	140
3.4.3	Can we accurately and cheaply determine which anthelmintics will be effective? .....	143
3.5	Conclusion .....	148
4	One site does not fit all: difficulties in identifying markers of IVM resistance using a reduced representation sequencing method .....	149
4.1	Abstract .....	149
4.2	Introduction .....	150
4.3	Results .....	154
4.3.1	Farm information and FECRT data .....	154
4.3.2	ddRAD-Seq read quantity and quality .....	155
4.3.3	Reads aligned well to the reference genome .....	158
4.3.4	Over 3 million RADloci identified, but very few shared between individuals.....	160
4.3.5	A diverse population but with the presence of contamination: kinship analysis .....	161
4.3.6	Many sites were out of Hardy-Weinberg Equilibrium .....	167
4.3.7	Six individuals were missing proportionally more common RADloci than others.....	168
4.3.8	Linkage disequilibrium decay was rapid.....	170

4.3.9	Low and variable SNP density across the genome .....	172
4.3.10	$F_{ST}$ analysis identifies genetically differentiated RADloci as potential markers of IVM resistance.....	174
4.3.11	RADlocus 587481 had the highest $F_{ST}$ value SNP, yet on closer inspection did not appear to be under IVM selection.....	182
4.3.12	Linkage disequilibrium across high $F_{ST}$ contigs.....	186
4.4	Discussion .....	187
4.4.1	Pre- and post-IVM sample populations appear as a single population	188
4.4.2	The sample size chosen was too small .....	188
4.4.3	Shared RADloci were far fewer than identified RADloci .....	189
4.4.3.1	Unequal amounts of DNA sequenced.....	190
4.4.3.2	Different RADloci may have been produced between individuals .....	191
4.4.4	Linkage disequilibrium decayed rapidly .....	193
4.5	Conclusion .....	194
5	A tale of two farms: validation of a Pool-Seq technique to investigate IVM resistance in UK field populations .....	195
5.1	Abstract .....	195
5.2	Introduction .....	195
5.3	Results .....	199
5.3.1	Farms used and FECRT data.....	199
5.3.2	A smaller percentage of reads aligned post-treatment compared with pre-treatment .....	200
5.3.3	Coverage varies between pre- and post-treatment samples .....	202
5.3.4	Sample pools are diverse, with some evidence of selection by IVM204	
5.3.5	Identification of differentiated genomic regions between pre- and post-treatment samples.....	213
5.3.6	Comparison with ddRAD-Seq results .....	223
5.4	Discussion .....	224
5.4.1	Whether a single locus, or multiple loci, contribute to IVM resistance on each farm is difficult to discern.....	225
5.4.2	Post-IVM samples are still highly diverse, but populations are less diverse than expected.....	227
5.4.3	Why might fewer reads have aligned post-treatment compared with pre-treatment? .....	228
5.4.4	Sample coverage was lower than expected - and may have affected sample size, increasing differences between technical replicates .....	228
5.5	Conclusion .....	230
6	Sifting the wheat from the chaff: Identifying genes which might be associated with IVM resistance .....	231
6.1	Abstract .....	231
6.2	Introduction .....	232

6.3	Results .....	233
6.3.1	Preliminary annotation of the Tci2 genome .....	233
6.3.2	Individual farm analysis identifies many genes of potential interest in IVM resistance .....	233
6.3.3	SNPs predicted to have high or moderate impact within regions of genetic differentiation .....	244
6.3.4	Comparison of results from both farms identified some common regions potentially under selection by IVM.....	249
6.3.5	Comparison of results with other GWAS.....	256
6.3.5.1	<i>T. circumcincta</i> New Zealand study .....	256
6.3.5.2	<i>H. contortus</i> studies .....	258
	Comparisons of <i>H. contortus</i> samples from across the globe .....	258
	<i>H. contortus</i> genetic cross (BUG Consortium) .....	258
6.3.5.3	<i>Onchocerca volvulus</i> .....	262
6.3.6	P-glycoproteins did not show copy number variation between pre- and post-treatment samples .....	264
6.3.7	Investigation of multiple gene copies in the Tci2 genome .....	271
6.3.8	Comparison of genes identified within each genome alignment analysis.....	275
6.4	Discussion .....	279
6.4.1	Neuronal genes.....	279
6.4.2	Metabolic and transporter genes .....	284
6.4.3	Regulatory genes and transposable elements .....	286
6.4.4	Multiple gene copies have potential to confound analysis .....	288
6.5	Conclusion .....	290
7	Discussion: Identifying the genetic basis of IVM resistance proved extremely challenging .....	291
7.1	Conclusion .....	303
	Appendices .....	305
	List of References .....	308



## List of Tables

Table 1.1: Anthelmintic classes licensed to treat <i>Teladorsagia circumcincta</i> in sheep in the UK. ....	26
Table 1.2: Tests described to detect anthelmintic resistance in ruminant strongyle nematodes. ....	32
Table 1.3: Prevalence of anthelmintic resistance on farms in the British Isles. .	34
Table 1.4: Mechanisms of resistance in parasitic strongyles. ....	49
Table 2.1: Primer sequences used to speciate strongyle larvae. ....	68
Table 2.2: Primer sequences for microsatellites. ....	71
Table 2.3: Primer sequences for pyrosequencing of the BZ resistance associated SNPs in $\beta$ -tubulin isotype-1. ....	73
Table 2.4: Primer sequences for <i>hprt-1</i> alleles. ....	74
Table 2.5: Primer sequences for core eukaryotic genes. ....	77
Table 3.1: Main management events on Farm 1. ....	103
Table 3.2: Anthelmintic treatments during 2016 on Farm 1. ....	105
Table 3.3: Farm 1 ewe strongyle faecal egg counts. ....	107
Table 3.4: Farm 1 lamb strongyle faecal egg counts. ....	109
Table 3.5: Farm 1 lamb <i>Nematodirus</i> faecal egg count. ....	111
Table 3.6: Benzimidazole and ivermectin faecal egg count reduction tests performed in September 2016 on Farm 1. ....	115
Table 3.7: Control wells of the egg hatch tests performed with samples taken during the faecal egg count reduction test ....	119
Table 3.8: Egg hatch test results. ....	119
Table 3.9: Control wells of the larval development test. ....	122
Table 3.10: Larval development test results. ....	122
Table 3.11: Microsatellite global AMOVA results as a weighted average over all loci. ....	132
Table 3.12: Individual microsatellite $F_{ST}$ values between samples. ....	132
Table 3.13: Sample pairwise $F_{ST}$ values for Farm 1. ....	134
Table 4.1: Faecal egg count reduction test results on Farm 2. ....	154
Table 4.2: Genes containing a RADlocus of interest identified following manual annotation of selected high $F_{ST}$ contigs. ....	176
Table 4.3: Genes within 10 kb of a RADlocus of interest identified following manual annotation of selected high $F_{ST}$ contigs. ....	177
Table 4.4: Genes within 50 kb of a RADlocus of interest identified following manual annotation of selected high $F_{ST}$ contigs. ....	178
Table 4.5: Genes within 75 kb of a RADlocus of interest identified following manual annotation of selected high $F_{ST}$ contigs. ....	181
Table 5.1: Faecal egg count reduction test results on Farm 2, 2013 and 2016. ....	200
Table 5.2: FECRT results on Farm 3, 2015. ....	200
Table 5.3: Summary of Pool-Seq alignment. ....	202
Table 5.4: Pool-Seq genome-wide mapped coverage per sample. ....	203
Table 5.5: Pool-Seq $F_{ST}$ thresholds. ....	218
Table 5.6: Counts of genomic regions and SNPs identified as being of interest on each of Farms 2 and 3. ....	223
Table 6.1: Numbers of Tci2 annotated genes and their predicted proteins within genetically differentiated regions. ....	235
Table 6.2: Farm 2 predicted homologues annotated with GO terms associated with neuronal tissues or processes. ....	237
Table 6.3: Farm 3 predicted homologues annotated with GO terms associated with neuronal tissues or processes. ....	239

Table 6.4: Farm 2 predicted homologues annotated with GO terms associated with lipid metabolism or storage.....	240
Table 6.5: Farm 3 predicted homologues annotated with GO terms associated with lipid metabolism or storage.....	242
Table 6.6: Farm 2 predicted homologues annotated with GO terms associated with female related terms. ....	243
Table 6.7: Farm 3 predicted homologues annotated with GO terms associated with female related terms. ....	243
Table 6.8: Farm 2 predicted homologues annotated with GO terms associated with male related terms.....	244
Table 6.9: Farm 3 predicted homologues annotated with GO terms associated with male related terms.....	244
Table 6.10: Farm 2 moderate impact SNPs predicted by SnpEff, which were present within genetically differentiated 10 kb windows. ....	246
Table 6.11: Farm 2 gene transcripts containing SNPs of moderate impact. ....	247
Table 6.12: Farm 2 gene transcripts putatively impacted by a SNP of moderate impact in a nearby gene.....	248
Table 6.13: Numbers of Tci2 annotated genes and predicted proteins within genetically differentiated regions which were identified in both farm analyses. ....	250
Table 6.14: 10 kb windows of the <i>Teladorsagia circumcincta</i> Tci2 draft genome identified on both Farm 2 and Farm 3 with $F_{ST}$ values greater than all comparison thresholds. ....	251
Table 6.15: Proteins predicted to be homologues or equivalent PRJNA72569 proteins of Tci2 genes located within separate regions of differentiation in each farm analyses. ....	254
Table 6.16: Genes shared between Farm 2 and Chromosome V peak of the BUG Consortium <i>Haemonchus contortus</i> genetic cross. ....	260
Table 6.17: Genes within genetically differentiated regions in Farm 2 data with potential homologues also identified in an <i>Onchocerca volvulus</i> ivermectin study. ....	263
Table 6.18: Genes within genetically differentiated regions in Farm 3 data with potential homologues also identified in an <i>Onchocerca volvulus</i> ivermectin study. ....	263
Table 6.19: Farm 2 <i>Caenorhabditis elegans</i> homologues with similar description to <i>Onchocerca volvulus</i> genes. ....	264
Table 6.20: Mean and median read counts per 5000 bp window for each sample. ....	266
Table 6.21: Genes and their predicted homologues/equivalent PRJNA72569 genes within or containing a region of altered copy number by read count. ....	267
Table 6.22: Consensus regions potentially under ivermectin selection identified on Farm 2 following alignment to both <i>Teladorsagia circumcincta</i> reference genomes. ....	276
Table 6.23: Consensus regions potentially under ivermectin selection identified on Farm 3 following alignment to both <i>Teladorsagia circumcincta</i> reference genomes. ....	278

## List of Figures

Figure 1.1: Diagram of the UK sheep industry. ....	19
Figure 2.1: Cuvette method for faecal egg counts. ....	63
Figure 2.2: Primer positions to amplify <i>hprt-1</i> gene copies. ....	75
Figure 2.3: Schematic representation of filtering of Farm 3 data. ....	89
Figure 2.4: Schematic diagram showing filtering of CNV regions. ....	93
Figure 3.1: Faecal egg count reduction test guidelines. ....	98
Figure 3.2: Strongyle species identification methods. ....	99
Figure 3.3: <i>Teladorsagia circumcincta</i> lifecycle ....	100
Figure 3.4: Strongyle faecal egg counts over time on Farm 1. ....	108
Figure 3.5: <i>Nematodirus</i> faecal egg counts over time on Farm 1. ....	110
Figure 3.6: Species proportions of cultured strongyle larvae on Farm 1 over time. ....	113
Figure 3.7: Farm 1 benzimidazole and ivermectin faecal egg count reduction test strongyle species proportions. ....	116
Figure 3.8: Farm 1 benzimidazole and ivermectin faecal egg count reduction test mean strongyle eggs per gram by species proportion. ....	117
Figure 3.9: Dose response curve of the egg hatch test performed using samples from the benzimidazole faecal egg count reduction test. ....	120
Figure 3.10: Dose response curve of the egg hatch test performed using the Pre-ivermectin faecal egg count reduction test sample. ....	121
Figure 3.11: Dose response curve of the larval development test performed using samples obtained during an ivermectin faecal egg count reduction test. ....	123
Figure 3.12: Dose response curve of the larval development test performed using the Pre-benzimidazole faecal egg count reduction test sample. ....	124
Figure 3.13: Lamb egg hatch test results over the season. ....	126
Figure 3.14: <i>Teladorsagia circumcincta</i> offspring genotypes determined by pyrosequencing before and after benzimidazole treatment of lambs. ....	128
Figure 3.15: <i>Teladorsagia circumcincta</i> offspring genotypes determined by pyrosequencing before and after ivermectin treatment of lambs. ....	129
Figure 3.16: Microsatellite allele counts. ....	132
Figure 3.17: Sample allele proportions for Farm 1 microsatellites. ....	133
Figure 3.18: Average soil and air temperature. ....	137
Figure 4.1: Ivermectin Faecal egg count reduction test Farm 2, 2016, with eggs per gram attributed by species. ....	155
Figure 4.2: Total number of read pairs sequenced per individual. ....	156
Figure 4.3: Identification of sequenced samples by whole genome amplification batch number. ....	158
Figure 4.4: BWA-MEM alignment percentages per individual. ....	160
Figure 4.5: Genotype plot of all individuals. ....	162
Figure 4.6: Genotype plot of all individuals ordered by treatment sample. ....	163
Figure 4.7: Genetic distance tree of all 62 ddRAD-Seq individuals. ....	164
Figure 4.8: PCA plot of all individuals. ....	165
Figure 4.9: ddRAD-Seq kinship analysis of all 62 individuals. ....	166
Figure 4.10: PCA plot of 60 individuals. ....	167
Figure 4.11: Percentage of shared variant sites missing per individual. ....	169
Figure 4.12: Percentage of variant sites missing per individual compared with sequences aligned. ....	170
Figure 4.13: Linkage decay plot. ....	172
Figure 4.14: ddRAD-Seq SNP density across the genome. ....	173
Figure 4.15: ddRAD-Seq SNP $F_{ST}$ Manhattan plot. ....	174

Figure 4.16: RADlocus 587481 position. ....	182
Figure 4.17: A Neighbour-Joining tree of the consensus fasta sequences of RADlocus 587481. ....	184
Figure 4.18: Contig 146 variant sites. ....	185
Figure 4.19: Contig mean $F_{ST}$ value compared with mean $r^2$ value. ....	187
Figure 5.1: Pool-Seq read GC content distribution. ....	201
Figure 5.2: Coverage depth site frequency. ....	203
Figure 5.3: Read count in 6145 bp windows along the genome. ....	204
Figure 5.4: Pearson correlation coefficients of genome-wide allele frequencies. .....	205
Figure 5.5: Tajima's $P_i$ by sample. ....	206
Figure 5.6: Tajima's $P_i$ by treatment time point. ....	207
Figure 5.7: Tajima's $D$ by samples. ....	209
Figure 5.8: Tajima's $D$ by treatment time point. ....	210
Figure 5.9: Farm 2 genome-wide Tajima's $D$ . ....	211
Figure 5.10: Farm 3 genome-wide Tajima's $D$ . ....	212
Figure 5.11: Farm 2 Manhattan plots comparing different distinct window sizes and fractional coverages. ....	214
Figure 5.12: Farm 3 Manhattan plots comparing different distinct window sizes and fractional coverages. ....	215
Figure 5.13: 10 kb window $F_{ST}$ pairwise comparisons. ....	216
Figure 5.14: SNP $F_{ST}$ pairwise comparisons. ....	217
Figure 5.15: Manhattan plots of Farm 2 and Farm 3 Pool-Seq data. ....	219
Figure 5.16: Farm 2 vs Farm 3 10 kb window $F_{ST}$ values. ....	221
Figure 6.1: Farm 2 copy number variation. ....	265
Figure 6.2: A Neighbour-Joining tree of P-glycoprotein amino acid sequences. ....	269
Figure 6.3: Alignment of amino acid sequences of Tci2 gene copies of PGP-3. ....	270
Figure 6.4: Pairwise alignment of the two copies of Tci2 $\beta$ -tubulin isotype-1 genomic sequence. ....	272
Figure 6.5: Alignment of HPRT-1 with orthologues. ....	274
Figure 6.6: Three Tci2 gene copies of <i>hprt-1</i> aligned. ....	274
Figure 7.1: Relative position of contigs within the <i>Teladorsagia circumcincta</i> genome. ....	298

## List of Accompanying Material

USB stick containing Appendices

## Acknowledgement

Thanks ever so much Eileen Devaney and Roz Laing for supervising me throughout my PhD! I have really enjoyed the last four years and have learnt a huge amount. Thanks to all the others in the lab who have been so supportive, especially Kirsty Maitland for teaching me how to do a PCR (and everything else) and Vicky Gillan for helping me understand my qPCRs. Thanks Steve Doyle for being so patient and cheerfully teaching me bioinformatics... again and again and again. Thanks James McGoldrick for teaching me how to ID strongyle larvae morphologically and for letting me dabble in real parasitology every so often. Thanks to Collette Britton and Antony Page for your help with understanding *C. elegans* genes and behaviour. Thanks as well Collette and Willie Weir for assessing me and providing lots of helpful input each year! Thanks to all those at the Moredun Research Institute, especially Ali Morrison, Dave Bartley, Leigh Devin and Lynsey Melville for all your help with FECs and bioassays - it was a lot of fun! Thanks also for helping to identify a suitable farm for sequencing and providing various lab isolates over the years. Thanks to Frank Turnbull and Lynsey for all your advice and for collecting the strongyle eggs for me from across the UK. Thanks to everyone at AHDB Beef and Lamb who were so friendly and inclusive when I visited for my placement! Special thanks to Mary Vickers and Lis King, my supervisors, and to Karen Morris who was fantastic at organising literally everything! Also special thanks to Sam Boon, Emma Steele, Liz Genever, Karen Fairlie-Clarke, Ben Strugnell and Fiona Lovatt for finding ways to enable me to work directly with sheep farmers and even do some farm animal pathology as part of the PhD! Thanks as well to James Cotton for all your help with bioinformatics and Nancy Holroyd for sequencing my worm DNA. Thanks to Neil Sargison and Kim Hamer for all your help with the field work, and for discussing ideas throughout the PhD. Thanks to Umer Chaudhry for your help and for patiently answering questions. Thanks to all those who worked on the BUG project - especially Katie Bull, Hannah Vineer and Eric Morgan for your advice, encouragement and for providing a suitable field population for sequencing! Thanks to all those who made CORE so interesting and to Nicola Beesley for the fun discussions. Thanks Taya, Steve and Brian for a good viva. Thanks to the other PhD students, to my wonderful family, friends and finally Jesus - without whom none of this would have been anywhere near as smooth or as enjoyable!!!

## **Author's Declaration**

The work presented within this Thesis was performed entirely by the author unless otherwise stated. The work described in the Thesis is unique and will not be submitted elsewhere for any other degree or qualification at any other university. Some of the work in Chapter 3 has been published as part of the paper McIntyre et al., 2018 (included in Appendices), or as part of Hamer et al., 2019 (referenced in the text). This work was funded by BBSRC (grant numbers BBN50385X/1 and BB/M003949/1), AHDB Beef and Lamb and the KTN (6130010011).

Jennifer R. McIntyre, November 2019

## Abbreviations

7TM-GPCR	7-transmembrane G-protein coupled receptor
AChR	Acetylcholine receptor (nAChR = nicotinic)
BCS	Body condition score
BLAST	Basic Local Alignment Search Tool (BLASTP = protein to protein, BLASTN = nucleotide to nucleotide, BLASTX = translated nucleotide to protein, TBLASTN = protein to translated nucleotide)
BQ	Base quality
BWA-MEM	Burrows Wheeler Aligner using super-Maximal Exact Matches
BZ	Benzimidazole
CET	Controlled efficacy test
CMH	Cochran-Mantel-Haenszel test
CP	Crude protein
ddNTP	Dideoxynucleotide triphosphate
ddRAD-Seq	Double digest Restriction-site associated DNA Sequencing
DMSO	Dimethyl sulfoxide
ED <sub>50</sub>	Effective dose at which 50% of individuals do not develop
EHT	Egg hatch test
epg	Eggs per gram (of faeces)
FEC	Faecal egg count
FECRT	Faecal egg count reduction test



F <sub>ST</sub>	Fixation index (how fixed alleles are between the sample and the total population)
GFP	Green fluorescent protein
GLM	General linear model
GluCl <sub>s</sub>	Glutamate-gated chloride channel subunits
GR	Good Responder
GWAS	Genome wide association study
HGVS	Human genome variation society notation
HPD	Highest Posterior Density interval
ITS2	Internal transcribed spacer region 2
IVM	Ivermectin
IVM-CRC	Ivermectin continuous release capsules
kb	Kilobase (used to denote basepair length)
kbp	Kilobasepair (used to denote genomic location)
L1, L2, L3, L4, L5	Larval stages of strongyle nematodes. L1-L3 are free-living, L3 is the infective stage. L4 and L5 are within the host.
LD	Linkage disequilibrium
LDT	Larval development test
LEV	Levamisole
LFIT	Larval feeding inhibition test
LGIC	Ligand-gated ion channels
LMIT	Larval migration inhibition test
MAPQ	Mapping quality

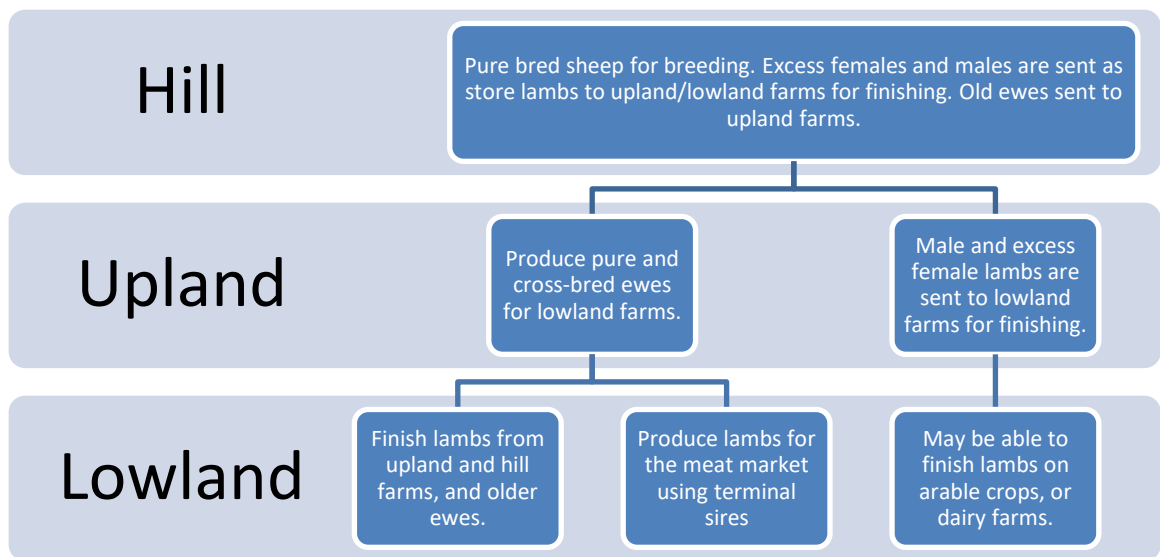
MHco	Moredun <i>Haemonchus contortus</i>
MPTL	Monepantel
MOX	Moxidectin
MRI	Moredun Research Institute, Edinburgh
MTci	Moredun <i>Teladorsagia circumcincta</i>
NGS	Next generation sequencing
QTL	Quantitative Trait Locus
QUAL	Quality
PCR	Polymerase chain reaction
PGE	Parasitic gastroenteritis
P-gp	P-glycoprotein
Pool-Seq	Pooled sequencing
PPR	Peri-parturient rise
RAD marker	SNP within a RAD-Tag
RAD-Seq	Restriction-site associated DNA sequencing
RAD-Tag	Region sequenced directly adjacent to a restriction site
RNA-Seq	RNA sequencing (transcriptome sequencing)
RR	Resistance ratio
SCOPS	Sustainable Control of Parasites in Sheep ( <a href="http://www.scops.org.uk">www.scops.org.uk</a> )
SD	Standard deviation
SNP	Single nucleotide polymorphism
SOR	Sub-optimal responder

SQP	Suitably qualified person
T <sub>A</sub>	Primer annealing temperature
TBZ	Thiabendazole
WGA	Whole genome amplification
WGS	Whole genome re-sequencing
WSI	Wellcome Sanger Institute, Hinxton

# 1 Introduction

## 1.1 Sheep Farming Systems

Sheep and goats are farmed worldwide, with New Zealand and Australia, in particular, being known for their high levels of sheep production. There is no one-size-fits-all approach, and sheep farming methods, breeds and outputs vary widely across the globe. In the UK, sheep farming is predominantly for meat, although in the past, wool was also a considerable production output, so much so that the Romney was exported from the UK to NZ, as a hardy, efficient dual-purpose breed (<http://romneysuk.com>). The UK sheep industry is divided into three main strata: hill, upland and lowland farms (Figure 1.1). Sheep move between and within these strata, enabling spread of parasites and anthelmintic resistance.



**Figure 1.1: Diagram of the UK sheep industry. In the UK, the sheep industry is stratified. Three separate farming 'regions' are present; hill, upland and lowland.**

The summer national flock in the UK consists of about 33.7 million sheep (DEFRA, 2018). Flocks can range in size from several thousand sheep down to only a handful, which may be kept as a hobby, or to conserve a rare breed. Generally speaking, flocks lamb in the late winter/early spring, and lambs are either finished for the meat market in late summer/early autumn, or sold for breeding. Those lambs which have failed to attain their market weight may be

sold as 'stores' and finished more slowly over winter in areas of the country which have sufficient pasture or arable crops. Some lambs may be retained by the breeder for a longer period of time before being sold. Traditionally, ewe lambs have been bred so as to lamb at two years of age, however, it is now encouraged to mate ewe lambs when they are about eight months of age, if they are ready, in a bid to reduce greenhouse gas emissions by up to 9.4% per kg of meat produced over their six-year lifetime (Rees and Phillips, 2016). This means that ewe lambs for breeding, which once were retained on their farm of origin until adulthood, may be increasingly sold when only five to six months old, at an age when they are unlikely to have developed robust immunity against parasites and may have high faecal egg counts (FECs) (Smith et al., 1985). Very few, if any, UK farms are truly closed, with almost all farmers buying in male breeding stock each year, predominantly in the autumn. Although most lambing occurs in the spring, the climatic conditions within the UK ensure that the coming of spring varies across the country, and as a result, lambing is usually earlier, sometimes by several months, in southern England compared to northern Scotland, adding yet another level of complexity to the overall management of the national flock.

Farmers generally try to aim for a lamb about 40 kg liveweight, as the consumer market is such that lamb at this weight achieves optimal price, with lambs under- or overweight potentially selling for less per kg than those within the specified weight bracket. In addition, lamb price varies week to week and month to month, so farmers must plan ahead to determine when they wish to lamb and how quickly they want to finish their lambs to sell. In this system, many external factors including both weather conditions and disease, can seriously affect the potential profit farmers make each year. The majority of farms receive subsidy payments from the government, and for many sheep farms these payments are the means of providing income for the farm over and above that required to break-even (AHDB, 2017). In the future, dependent on trade deals with Europe and the rest of the world, and subsidies following BREXIT, the economics of farming in the UK may change dramatically and it is possible that many farmers will struggle to stay in business (AHDB, 2017). It is therefore essential that production-limiting diseases are managed effectively with a view to reducing

their impact, both in terms of management costs and detrimental effect on sheep production.

## 1.2 Parasitic Gastroenteritis

The Agricultural and Horticultural Development Board (AHDB) Beef and Lamb estimated that parasitic gastroenteritis (PGE), a severe production limiting disease, reduced the price per lamb by £10 (Lovatt and Stubbings, 2018). This is a considerable amount given that the price per 40 kg lamb averaged £74.84 in 2017, with store lambs (those finished more slowly overwinter), reaching only £54.15 on average (AHDB, 2018). These costs include the increased feed consumed by a lamb to reach the same final weight, the cost of treatment (anthelmintics), increased labour required to raise a lamb and reduced availability of pasture for other purposes (Lovatt and Stubbings, 2018). PGE has consistently been the commonest cause of death diagnosed by the Animal and Plant Health Agency (APHA) in both sheep and goats (APHA, 2018). Although the presentation varies, PGE is considered ubiquitous in sheep farming, and is a worldwide concern of farmers (Kaplan and Vidyashankar, 2012). With both clinical and sub-clinical presentations, PGE manifests with various clinical signs, including: poor weight gain, anorexia, diarrhoea, low milk yield, death and pallor (Durham and Elliott, 1975). Pathology includes gastroenteritis, with the recruitment of mast cells, eosinophils, neutrophils and lymphocytes to the mucosal and submucosal layers (Durham and Elliott, 1975; Balic et al., 2003). Dehydration, hypoproteinaemia, poor digestion of the diet and anaemia are evident (Baker et al., 1959; Lawton et al., 1996). PGE is caused by a wide range of parasites including: *Nematodirus battus*, predominantly in the spring in the UK, in young lambs, which can manifest as severe sudden onset diarrhoea and/or death; *Trichostrongylus spp.* cause most disease (known as ‘black scours’) during the autumn and winter months; *Haemonchus contortus*, responsible for Haemonchosis, is haematophagous, and although diarrhoea is not usually a clinical sign, anaemia is common and can be fatal; and *Teladorsagia circumcincta*, the ‘brown stomach worm’, responsible for Teladorsagiosis, and the commonest cause of PGE in growing lambs over the summer in the UK (Burgess et al., 2012; Mitchell, 2016). Other nematode species are also potentially involved in small ruminant PGE, however they are generally considered less pathogenic than those mentioned above and usually contribute to, rather than primarily cause, PGE.

Each of these strongyle species have slight differences in their lifecycles affecting the timing, presentation, management and severity of PGE. Farm management practices can impact upon which species survive and thrive on a holding. This is perhaps most evident when considering the differences seen in the abundance of species in unmanaged wild flocks (Wimmer et al., 2004; Sinclair et al., 2016), and the comparative lack of diversity noted on some commercially managed farms (Redman et al., 2015).

### 1.3 Teladorsagiosis

*T. circumcincta* is an abomasal trichostrongylid nematode of sheep and goats. Previously known as *Ostertagia circumcincta*, it was thought to be related to the cattle parasite *Ostertagia ostertagi*. However, it was determined to belong to a separate genus in the 1980s by morphological differences of the lateral rays in the bursa and an absence of the proconus of the male (Gibbons and Khalil, 1982; Lichtenfels et al., 1988; Lichtenfels and Hoberg, 1993). Following ingestion of the infective L3 stage from pasture the larva exsheaths within the rumen (Bekelaar et al., 2018), passes to the abomasum and enters the gastric glands of the abomasal wall. There it causes a severe inflammatory reaction, provoking a small raised lesion surrounding the gland and grossly visible on examination, providing the classic 'Moroccan leather' appearance (McKellar, 1993). It undergoes two moults to become an immature adult, and emerges to complete development on the mucosal surface of the abomasal wall. Once fully developed it mates, producing eggs, roughly 100-200 per female per day (Cole, 1986). The pre-patent period from ingestion to first appearance of eggs in the faeces is about 17-21 days, but can be shorter with one study recording eggs in the faeces of infected sheep at just 14 days post-infection (Barrett et al., 1998). As a consequence of infection, the serum pepsinogen and gastrin increase, and in the initial stages pre-patency, the abomasal pH increases, although it returns to normal levels once patency is established in experimentally infected sheep (Lawton et al., 1996).

*T. circumcincta* is ubiquitous on UK sheep farms, being present on all farms tested in Scotland, England, Wales and Northern Ireland in a recent study (Burgess et al., 2012), and in several independent studies of farms in Wales (Hybu Cig Cymru, 2015), Northern Ireland (McMahon et al., 2013b), Scotland

(Bartley et al., 2003), and southeast England (Glover et al., 2017). Although it is found worldwide (Kaplan, 2004), *T. circumcincta* is predominantly a parasite of small ruminants in temperate climates (O'Connor et al., 2006).

*T. circumcincta* is a particular problem for UK farmers as it is expertly adapted to both the sheep host and the farming year. The parasite is present within all sheep on a holding, but the egg output attributable to *T. circumcincta* is very low in non-pregnant adult sheep. Nevertheless, in late pregnancy immunity against *T. circumcincta*, and other strongyles (excluding *N. battus*) is reduced and egg output from all strongyles, but especially *T. circumcincta*, increases dramatically (Gibson, 1973; Jansen, 1987). This phenomenon is known as the peri-parturient rise (PPR) and is key to the maintenance of disease-inducing and production-limiting levels of *T. circumcincta* on farm (Coop et al., 1982). Following lambing, the PPR continues for six weeks, before the ewe's immune system reasserts itself leading to a reduction in egg output (Gibson, 1973). Lambs initially eat very little grass, but over time begin to ingest greater quantities, increasing their intake of L3 from the pasture. Due to the nature of the PPR, and the time taken for eggs to develop to L3 in the UK spring, pasture infectivity is usually high at the time at which lambs begin feeding on grass in earnest (Gibbs, 1986b). Lambs take between four to six months to develop immunity to *T. circumcincta* (Smith et al., 1985), and during this time they greatly multiply the number of larvae residing on pasture. This lack of immunity towards parasites causing PGE severely impacts on lamb growth rate if PGE is not appropriately managed within the flock (Coop et al., 1982). Unlike most other species, *T. circumcincta* is able to develop at very low temperatures (Salih and Grainger, 1982), and can survive on pasture overwinter, with infective L3 still found as late as 16-18 months following egg deposition (Gibson, 1973). These overwintered larvae also contribute to the PPR, as they are ingested by ewes in early lactation, and thus contribute to the subsequent pasture population during the later spring and summer (Sargison et al., 2002). *T. circumcincta* is capable of hypobiosis, arresting development overwinter within the host, ensuring both that adults emerge when ewes experience low immunity and are less effective at suppressing fecundity of parasites, and also when eggs are more likely to successfully develop to L3 than at the start of winter (Gibbs, 1986b).



In the faeces, *T. circumcincta* develops from an egg to an infective L3, taking between 14 days to 10 weeks dependent on weather conditions (Gibson and Everett, 1972). The L3 migrate onto pasture in response to moisture (van Dijk and Morgan, 2011). Nevertheless, unlike other strongyle parasites, such as *H. contortus* and *Trichostrongylus colubriformis*, this migration occurs slowly over many months, as a trickle, rather than *en mass* (Gibson and Everett, 1972; Wang et al., 2014). This allows the maintenance of a sub-population within the relative safety of the faeces, ensuring pasture levels of infectivity remain higher for longer (Gibson and Everett, 1972). L3 are able to survive on pasture for a year or longer, but are more exposed and are prone to desiccation (Gibson and Everett, 1972; van Dijk et al., 2009). In addition, the survival of eggs developing to L3 appears linked to the time of year at which they are shed, with those laid in winter providing lower yields of L3, and these only surviving for a maximum of five months on pasture in a UK study (Gibson and Everett, 1972).

## 1.4 Management of Parasitic Gastroenteritis

Management options for PGE vary, and not all practices will be available to all farmers. Almost all farmers rely to a greater or lesser extent on anthelmintics (Burgess et al., 2012). These drugs are a diverse set of products, which are available without the requirement for veterinary input as they can all be obtained via a suitably qualified person (an SQP). SQPs are legally designated animal medicine advisors, who have been trained in the prescription or supply of certain medications, but are not necessarily medically trained. There are five anthelmintic classes which can be used to treat *T. circumcincta* in sheep in the UK (Table 1.1) (Abbott et al., 2012). Briefly, the earliest modern anthelmintic class introduced in 1961 was the benzimidazole class (1-BZ), with the product thiabendazole (Gordon, 1961). This compound, although still widely used in research, is no longer used on farm, but there are several other BZ anthelmintics licensed in the UK for sheep, namely: albendazole, fenbendazole, mebendazole, oxfendazole and ricobendazole (SCOPS, 2019). The second class of anthelmintic is the imidazothiazoles (2-LV), of which only the anthelmintic levamisole (LEV) is licensed for use in UK sheep. The third class are the macrocyclic lactones (3-ML), which consist of two sub-classes: avermectins and milbemycin oximes. Avermectins include ivermectin (IVM), doramectin, eprinomectin and abamectin, the latter only sold as a combined product with derquantel. IVM is increasingly

used by sheep farmers in the UK specifically to treat PGE in lambs, with moxidectin (MOX), a milbemycin oxime, also commonly used at the time of the PPR (Burgess et al., 2012). MOX is marketed in both 0.1% oral and 1% injectable formulations with five weeks persistency against re-infection with *T. circumcincta*, and a 2% injectable formulation with over 13 weeks persistency (Kerboeuf et al., 1995; Abbott et al., 2012). Macrocyclic lactones are also used to treat the ectoparasite *Psoroptes ovis*, both prophylactically and reactively, simultaneously affecting the strongyle population within the host (Doherty et al., 2018). MOX was introduced onto the UK market in 1991 but no new anthelmintics were developed from that time until 2009, when monepantel (MPTL) (an amino-acetonitrile derivative, class 4-AD) was released onto the market in NZ and Australia and a year later in the UK (Hosking et al., 2009; Bartley et al., 2019). In 2011, derquantel (a spiroindole, class 5-SI) was developed, and released as a combined product with abamectin (Little et al., 2011). Two final anthelmintics available to farmers are also flukicides (active against *Fasciola hepatica*): closantel and nitroxynil, both of which have narrow spectrum efficacy against *H. contortus* (Abbott et al., 2012). Farmers vary in their use of anthelmintics, both in frequency of dosing and choice of anthelmintic used for sheep by age, and time of year (Burgess et al., 2012). In the UK, the presence of *N. battus* requires many farmers to treat their lambs monthly in mid- to late-spring and the BZ anthelmintics are recommended for this (Abbott et al., 2012). However, recent reports indicate that resistance is developing and, furthermore, that many farmers choose to use other anthelmintic classes for nematodiosis treatments (Morrison et al., 2014; McMahon et al., 2017).

**Table 1.1: Anthelmintic classes licensed to treat *Teladorsagia circumcincta* in sheep in the UK.**

Anthelmintic class	Anthelmintics licensed in the UK for sheep	Date first introduced	First reports of resistance
1-BZ (benzimidazoles, ‘white drenches’)	Albendazole, fenbendazole, mebendazole, oxfendazole and ricobendazole	1961 (thiabendazole, no longer used to treat sheep) (Gordon, 1961)	(Drudge et al., 1964; Smeal et al., 1968; Britt, 1982)
2-LEV (imidazothiazoles, ‘yellow drench’)	Levamisole	c.1970 (Thienpont et al., 1966; Kaplan, 2004)	(Sangster et al., 1979; Hong et al., 1994, 1996)
3-ML (macrocyclic lactones, ‘clear drenches’, also injectable forms); sub-class avermectins	Ivermectin, doramectin, eprinomectin and abamectin	1981(Chabala et al., 1980; Campbell et al., 1983)	(van Wyk and Malan, 1988; Jackson et al., 1992a)
3-ML (macrocyclic lactones, ‘clear drenches’, also injectable forms); sub-class milbemycin oximes	Moxidectin	1991 (Kerboeuf et al., 1995)	(Sargison et al., 2005; Sargison et al., 2010)
4-AD (amino-acetonitrile derivative, orange)	Monepantel	2009 (2010 in UK) (Hosking et al., 2009; Kaminsky et al., 2009; Abbott et al., 2012)	(Scott et al., 2013; Van den Brom et al., 2015; Hamer et al., 2018)
5-SI (spiroindole, purple)	Derquantel (only available as a combination product with abamectin)	2011 (Little et al., 2011)	(Lamb et al., 2017) <sup>1</sup>

The anthelmintics available within each class are shown, and an indication of the first introduction of the anthelmintic class into the UK is provided, with early reports of anthelmintic resistance both worldwide and in the UK. <sup>1</sup>This may be due to the presence of *Haemonchus contortus* L4 achieving patency, rather than adult resistance and the authors advised further follow-up work was required to fully diagnose resistance.

Advice on management of PGE is provided to farmers by the knowledge transfer group, 'Sustainable Control of Parasites in Sheep' (SCOPS) and the levy board AHDB Beef and Lamb, (in addition to other levy boards, organisations and vets). Research has assessed both the uptake and effect of the SCOPS recommendations. SCOPS guidelines are followed by some farmers, however a significant proportion of farmers still use traditional practices (McMahon et al., 2013c). One study compared the effect of using SCOPS guidelines to treat sheep with anthelmintics, to traditional treatment management practices. No significant difference was noted in worm burden or lamb weight gain over three years between the farms monitored, however SCOPS farmers used significantly less anthelmintic than traditional farmers (Learmount et al., 2016b). Other than the use of anthelmintics, management options include; (i) using cattle as biological 'hoovers', effectively reducing the pasture larval contamination of sheep specific strongyle nematodes, including *T. circumcincta*. Cattle could be used in rotation with sheep, or alongside sheep, which may have an added benefit of increasing the overall nematode diversity (Giudici et al., 1999). (ii) Using arable crop rotation to effectively clean up a pasture, which can be subsequently re-seeded with grass. (iii) Alternatively, silage production, several cuts likely being better than a single cut, can be used. (iv) Growing certain forage crops, such as chicory, which may reduce the ability of larvae to develop on pasture (Marley et al., 2003), and (v) using nematophagus fungi, although this varies in its efficacy (da Silveira et al., 2017). In addition, copper has been investigated as a means of reducing parasite burden in sheep infected with *H. contortus* (Knox, 2002; Leal et al., 2014), and lastly, vaccination against nematode parasites is becoming an increasing possibility. Already, an *H. contortus* vaccine based on nematode gut proteases is in use in Australia and South Africa (Smith et al., 2001), although it is not available in the UK. A vaccine for a cattle pulmonary nematode, *Dictyocaulus viviparous*, based on irradiated L3, has been available and in use since the 1960s (Jarrett et al., 1960; Strube et al., 2015). Ongoing work at the Moredun Research Institute (MRI) aims to develop a recombinant subunit vaccine for *T. circumcincta* (Nisbet et al., 2016a) and *H. contortus* (Nisbet et al., 2016b; Sallé et al., 2018). While it is not possible to eliminate parasites from sheep flocks, it is possible to reduce the infective larval burden, or breed resilient or nematode-resistant animals reducing the impact of PGE on production losses (Fairlie-Clarke et al., 2019). Of

key importance in managing anthelmintics wisely is the maintenance of a susceptible population of worms *in refugia* (van Wyk, 2001), within which any resistant alleles will be diluted, prolonging the effectiveness of an anthelmintic in a flock. This population will be, by definition, unexposed to the anthelmintic and will include pre-parasitic stages on pasture, worms in untreated animals (both domestic and wild) and lifecycle stages within a treated animal, which are not affected by the anthelmintic. However, all classes of modern anthelmintics for sheep are active against both immature and adult strongyles, with many also capable of killing inhibited larval stages within the host. The population *in refugia* will not include free-living stages affected by anthelmintic, which reside within faeces containing anthelmintics. Nor will it apply in animals which are untreated by the farmer but which obtain anthelmintics through other means; for example by consuming milk from treated animals (Dever and Kahn, 2015), or licking anthelmintic from animals treated using a pour-on product (Laffont et al., 2001).

## 1.5 Anthelmintic Resistance

Resistance is defined as the ability of an individual to survive concentrations of an anthelmintic which are higher than that which might be expected to be therapeutic, or which some individuals within a normal, heterogeneous population, with natural tolerance towards the anthelmintic might be expected to survive (Le Jambre et al., 1976). On farm, and also in many research studies, anthelmintic efficacy is tested using a faecal egg count reduction test (FECRT) (Coles et al., 2006). This involves the collection of faeces from a group of animals both before and after treatment with an anthelmintic. FECs are performed at each sampling time point and the percentage reduction in egg output is calculated (Coles et al., 2006). For older anthelmintics, efficacy was not expected to be 100%, however, modern anthelmintics can provide close to 100% efficacy against strongyles (Wood et al., 1995). Anthelmintic resistance is thus generally considered when tests show that the anthelmintic is less than 95% effective, with the 95% lower confidence interval calculated at 90% reduction in egg output or less (Coles et al., 1992). If only one of these two conditions is met, resistance is suspected but cannot be confirmed (Coles et al., 1992).

Other means of determining anthelmintic efficacy are also used in research, with the gold standard being a Controlled Efficacy Test (CET) (Wood et al., 1995). To test a field population using a CET, faeces are collected from the farm, and L3 cultured and harvested. Worm-free sheep are then infected with a set number of these L3, and following establishment of patency, are treated with the anthelmintic under consideration. A FECRT is performed and the sheep are sacrificed, following which the number of adults surviving treatment are identified to species level, quantified and compared with an untreated control sheep (Wood et al., 1995; Bartley et al., 2019).

## 1.6 *In vitro* tests of anthelmintic sensitivity

*In vitro* bioassays allow determination of the effect of an anthelmintic on the free-living stages (Table 1.2). Perhaps the commonest assay in active use is the Egg Hatch Test (EHT), used primarily to determine BZ efficacy by the prevention of development to, and hatching of, L1 larvae (Le Jambre, 1976). This can also be used to determine LEV efficacy (Dobson et al., 1986), and for MPTL (Bartley et al., 2016), however it cannot be used to determine the efficacy of IVM as this is not ovicidal except at extremely high concentrations (Patel, 1997). Eggs must be harvested and the EHT set up within 3 hours of collection (or stored anaerobically), otherwise results are unreliable (Coles et al., 2006). Culture of larvae from eggs through to L3 over the course of a week can also be performed, as a Larval Development Test (LDT), with the anthelmintic (BZ, LEV or a ML), added at either the start of the test, or after 24 hours (Taylor, 1990). Several variations on the LDT set-up are possible, and variation is also apparent in the development of different strongyle species within the assay (Hubert and Kerboeuf, 1992; Gill et al., 1995; Heim et al., 2015). LDTs can be liquid based (Varady et al., 1996) or agar based (Gill et al., 1995). Each has different benefits and drawbacks, and neither test type appears to work well for all species.

In the presence of genetically modified *E. coli* producing GFP, the ability of LEV or MLs to inhibit pharyngeal pumping, and thus feeding, can be determined, known as a Larval Feeding Inhibition Test (LFIT). In practice, this requires harvesting and hatching eggs, incubating L1 larvae in varying concentrations of either LEV or IVM for two hours before further incubation with GFP labelled *E. coli* for an additional two hours (Alvarez-Sanchez et al., 2005b). The assay

performs better for IVM than for LEV, with resistance ratios (RR) of approximately 7.5 for both IVM resistant *H. contortus* and *T. circumcincta* isolates tested, compared to ratios of only 2.7 for LEV resistant *H. contortus* compared to susceptible isolates. Importantly, the sensitivity of L1 to IVM or LEV appears to vary over the time-course of an infection, and so it would be important to age match both the resistant and sensitive strains whenever the assay was performed, which may limit its application in the field (Alvarez-Sanchez et al., 2005b).

Using the infective L3 stage, the effect of IVM on somatic musculature can be determined using a Larval Migration Inhibition Test (LMIT) (Demeler et al., 2010a). Briefly, following incubation in the anthelmintic, larvae are tested for their ability to migrate vertically down through a fine mesh sieve over time (Demeler et al., 2010b). However, a more recent study assessing both motility (using the 'Worminator', a video capture technology (Storey et al., 2014)) and migration of field isolates of *H. contortus* and *Cooperia oncophora* found both tests to be unhelpful in determining resistance status to avermectins (George et al., 2017).

A common issue with all of the above-mentioned assays is the presence of mixed species populations, as between-species differences may impact upon assay interpretation. These include differences in development time and in natural tolerance of the species to the anthelmintic and other reagents used in the assay. This confounding effect is of concern in field populations, which often contain a mix of strongyle species (Boag and Thomas, 1977; Redman et al., 2015). Further complications arise when one species within a population is sensitive to an anthelmintic and another is resistant (Calvete et al., 2014). In addition, differences are apparent between different lifecycle stages (Sangster et al., 1988; Sangster and Bjorn, 1995; Varady and Corba, 1999), age of L3 (Geerts et al., 1989) and throughout the patency period of an infection (Varady and Corba, 1999) even within a single species. Within a field sample collected at any given point in time, it is expected that eggs will have been shed by parasites of different ages, as sheep continually graze infective pasture. Lastly, it has been shown that the ability to detect BZ resistance in a population using either a

FECRT or an EHT lacks sensitivity until at least 25% of the population is resistant (Martin et al., 1989).



**Table 1.2: Tests described to detect anthelmintic resistance in ruminant strongyle nematodes.**

Test	1-BZ	2-LEV	3-ML	4-AD	5-SI
CET	(Wood et al., 1995)	(Wood et al., 1995)	(Wood et al., 1995)	(Wood et al., 1995)	(Wood et al., 1995)
FECRT	(Coles et al., 2006)	(Coles et al., 2006)	(Coles et al., 2006)	(Raza et al., 2016c)	(Geurden et al., 2012)
EHT	(Le Jambre, 1976; von Samson-Himmelstjerna et al., 2009a)	(Dobson et al., 1986)		(Bartley et al., 2016)	
LDT	(Kotze et al., 2012)	(Hubert and Kerboeuf, 1992)	(Gill et al., 1995)	(Raza et al., 2016c)	
LMIT			(Demeler et al., 2010a)		
LFIT			(Alvarez-Sanchez et al., 2005b)		
Allele-specific PCR	(Silvestre and Humbert, 2000; Rufener et al., 2009a)				
Allele-specific qPCR	(Alvarez-Sanchez et al., 2005a)				
Pyrosequencing	(Skuce et al., 2010)				

An indication of whether a particular test is available for each of the five anthelmintic classes is provided. *In vivo* tests: CET – Controlled efficacy test, FECRT – Faecal egg count reduction test. *In vitro* tests: EHT – Egg hatch test, LDT – Larval development test, LMIT – Larval migration inhibition test, LFIT – Larval feeding inhibition test. Molecular tests: Allele-specific PCR, Allele-specific qPCR and Pyrosequencing.

## 1.7 Prevalence of Anthelmintic Resistance

The reported prevalence of anthelmintic resistance varies between studies and is likely partially dependent on the method used to detect it. In some parts of the world, particularly in South America, anthelmintic resistance is highly prevalent, to the extent that many farms have been diagnosed with multiple anthelmintic resistance (Kaplan and Vidyashankar, 2012). Within Europe, prevalence of anthelmintic resistance appears to vary between countries and host species (Rose et al., 2015). In 2012, a UK study surveyed farmers, who were asked whether resistance had been diagnosed in their flock by a veterinary surgeon (Burgess et al., 2012). Farmers reported known resistance to BZ and LEV anthelmintics only. However, both previous and subsequent studies within the UK and Ireland have identified larger numbers of farms with resistance to BZ, LEV and the macrocyclic lactones than were reported in that study (Table 1.3). Not all studies have diagnosed resistance to species level, and this likely has an impact on the interpretation of results reported (McMahon et al., 2013b). The time of year at which a study is performed also impacts on the prevalence detected (Hybu Cig Cymru, 2015), probably mostly due to species composition, although host age, and stage of infection may also play a role. Multi-drug resistant populations have been increasingly reported, following the first European diagnosis in Scotland in 2001 (Sargison et al., 2001; Sargison et al., 2010; Hybu Cig Cymru, 2015). Worryingly, few farmers appear to regularly test for anthelmintic efficacy (Morgan et al., 2012). Nevertheless, the use of macrocyclic lactones is increasing, which suggests a perceived lack of efficacy of older anthelmintics more commonly used previously, or a general increase in treatment frequency (Keegan et al., 2017b). Some will also be choosing to use macrocyclic lactones to treat *P. ovis* more frequently than previously, or may be buying combination flukicide and nematocide products, which can contain macrocyclic lactones.

**Table 1.3: Prevalence of anthelmintic resistance on farms in the British Isles.**

Study	Location	Prevalence (percentage) of Anthelmintic Resistance (Total number of farms tested)					Method used
		BZ	LEV	IVM	MOX	MON	
(Cawthorne and Cheong, 1984)	SE England	13.5% (52)	0% (52)	ND	ND	ND	FECRT <sup>1</sup>
(Grimshaw et al., 1994)	S England	100% (5)	40% (5)	0% (5)	ND	ND	FECRT
(Hong et al., 1996)	England and Wales	32.6% (138)	1 farm (total tested unclear)	0% (26)	ND	ND	FECRT/EHT/LDT/ Lab sheep infected <sup>1</sup>
(Bartley et al., 2003)	Scotland	64.4% (90)	ND	ND	ND	ND	EHT
(Bartley et al., 2006)	Scotland	ND	ND	35.3% (17)	ND	ND	FECRT
(Fraser et al., 2006)	SW England	28% (90) <sup>2</sup>	?	?	?	ND	Survey
(Mitchell et al., 2010)	Wales	56.3% (122)	36% (122)	ND	ND	ND	LDT
(Burgess et al., 2012)	UK	17.5% (118)	3.4%(118)	ND	ND	ND	Survey
(Good et al., 2012)	Ireland	88.2% (17)	38.9% (18)	ND <sup>3</sup>	ND	ND	FECRT
(McMahon et al., 2013b)	N Ireland	81% (26)	14% (7)	50% (14)	62% (21)	0% (3)	FECRT <sup>4</sup>
(Keane et al., 2014)	Ireland	70% (155)	48% (82)	24% (132)		ND	FECRT
(Hybu Cig Cymru, 2015)	Wales	94% (47)	68% (47)	51% (47)	19% (47)	ND	FECRT
(Kenyon et al., 2016)	S Scotland	100% (4)	75% (4)	100% (4)	100% (1)	ND	FECRT (Mini-FLOTAC)
(Glover et al., 2017)	SW England	96% (25)	60% (25)	67% (27)	ND	ND	FECRT
(Keegan et al., 2017b)	Ireland	68.5% (550) <sup>5</sup>	48.1% (316) <sup>5</sup>	37.5% (405) <sup>5,6</sup>	16% (163) <sup>5</sup>	ND	FECRT

For each study included, data are shown in terms of the percentage of farms with detectable anthelmintic resistance, as defined by the study authors. For each anthelmintic, the percentage of farms with resistance is provided, in addition to the total number of farms tested or surveyed in the study. The method used to determine prevalence is provided. In some studies multiple methods were used, and not all are provided here. <sup>1</sup>Note post-drench check at day 7, potential for under-estimation. <sup>2</sup>Two tests used combination products. <sup>3</sup>Suspected by LDT. <sup>4</sup>Excluded combination product results. <sup>5</sup>Number of tests, not farms. <sup>6</sup>Avermectins. ND = Not Done.

## 1.8 Ivermectin Resistance

IVM is a macrocyclic lactone, released onto the market in 1981 after being discovered by Omura and Campbell (Campbell et al., 1983). IVM induces a flaccid paralysis (Arena et al., 1995), leading to worm expulsion from the host. One of the most common anthelmintics in use by farmers, IVM is increasingly used in both the UK and Ireland (Burgess et al., 2012; Keane et al., 2014). There is a considerable prevalence of IVM resistance worldwide (Kaplan and Vidyashankar, 2012), and in the UK, over 50% of farms tested had detectable IVM resistance in recent studies (McMahon et al., 2013b; Hybu Cig Cymru, 2015).

### 1.8.1 IVM resistance is inherited as a fully or partially dominant trait

Several studies have investigated the aetiology of IVM resistance using genetic crosses between sensitive and resistant isolates to produce heterozygous offspring (F1 generation). One study suggested that, for *T. circumcincta* developing larvae, IVM resistance is partially dominant (Sutherland et al., 2003b). A multi-drug resistant (BZ, LEV, IVM, MOX) field isolate was exposed to intense selection pressure with BZ, LEV and IVM. The original population when tested was less resistant to IVM than the highly selected resistant strain, but equally resistant to MOX (Sutherland et al., 2003b). Females of this strain were then crossed with males of a sensitive isolate. IVM-continuous release capsules (CRC), which release a low concentration of IVM, prevented establishment of L3 of the sensitive isolate. In contrast, the selected resistant strain was able to establish infection. The F1 generation was capable of partial establishment of infection, based on FEC and worm burden data. Although this may suggest that IVM resistance in *T. circumcincta* is partially dominant, as the authors inferred, the potential for side-resistance to IVM conferred by MOX resistance mechanisms cannot be ruled out. Side-resistance occurs when resistance mutations associated with an anthelmintic confer resistance to other anthelmintics in the same class (Sangster, 1999). In addition, the resistant strain could contain heterozygous individuals. If so, there could be an increased proportion of homozygous susceptible individuals within the F1 generation compared to offspring derived from the resistant strain alone. The authors also used the same parasite strains to infect sheep subsequently treated with IVM. IVM resistance in

adult *T. circumcincta* appeared to be a dominant trait based on FEC and worm burden (Sutherland et al., 2002).

A separate study investigating the inheritance of IVM resistance in *H. contortus*, established a laboratory cross between an IVM resistant CAVR strain (CAVRS) and an IVM sensitive isolate (VRSG) (Le Jambre et al., 2000). The authors found that IVM resistance appeared to be autosomal and dominant in larvae (based on a LDT using avermectin B<sub>2</sub>). There was no difference in larval development *in vitro* in the presence of IVM between the F1 generation (offspring of the cross) and larvae of the resistant parental isolate (CAVRS) (Le Jambre et al., 2000). This was observed whether males or females of the resistant isolate were used in the cross. Nevertheless, following infection of sheep with the F1 generation of the cross, and *in vivo* treatment with IVM, 50% fewer males than females were found at post-mortem, suggesting some advantage of adult females exposed to IVM, and potentially a difference in resistance mechanisms between sexes (Le Jambre et al., 2000). More recently, a laboratory cross between males of the BZ, LEV and IVM resistant *H. contortus* isolate, MHco18(UGA), and females of the sensitive isolate, MHco3(ISE), found IVM resistance to be semi-dominant. A FECRT performed on a sheep infected with the F1 generation of the cross found that the adults were more resistant than MHco3(ISE), although less so than MHco18(UGA) (Laing, *pers comm*).

### **1.8.2 IVM affects motility, fecundity and reproduction**

IVM is known to affect the motility of nematodes when applied in micromolar concentrations (LMIT) (Demeler et al., 2010a), and to alter feeding behaviour by acting on the pharynx (LFIT, LDT) when free-living stages of nematodes are exposed to nanomolar concentrations (Gill et al., 1995; Bartley et al., 2009). The latter causes cessation of pharyngeal pumping and can induce starvation (Laing et al., 2012).

Evidence from field studies indicates that *in vivo* treatment with IVM also affects the reproductive system. Significantly increased mean numbers of eggs within adult *H. contortus* females were noted in sheep infected with an IVM resistant population at seven days post-IVM treatment compared with pre-treatment adults, and simultaneously lower FEC (Scott et al., 1991), suggestive of egg

retention. This followed an initial reduction in eggs within adult females at 48 hours post-treatment. This could indicate that IVM may have both an embryostatic effect and, in addition, affect egg expulsion in *H. contortus*. This could be by separate modes of action. However, it should be noted that this study only evaluated ten *H. contortus* females at each sampling time point (via an abomasal cannula), and the reduced FEC was not followed up with an overall worm count post-treatment. In addition to suppression of egg output by IVM, eggs collected post-IVM treatment from adult females were also found to larvate within 48 hours at 27 °C but to remain unhatched (Scott et al., 1991). Although hatching appeared delayed in the study by Scott et al. (1991), a second study found that IVM resistant *H. contortus* eggs were capable of developing to L3 providing they were first removed from faeces containing IVM (Tyrrell et al., 2002). A similar experiment using *T. circumcincta* infection of seven month old lambs dosed with IVM-CRC found that the number of eggs per adult female was significantly higher in those treated with IVM than in control sheep (Sutherland et al., 2003a). However there was no overall difference in the FEC (although adult worm burden was still reduced, on average, compared to the control, but not significantly). Neither was any difference identified in the ability of eggs to hatch or develop to L3 in faeces (Sutherland et al., 2003a). IVM has also been shown to reduce fecundity by an unknown embryostatic mechanism in the filarial nematode *Onchocerca volvulus* (Duke et al., 1991) and in mosquitoes (Deus et al., 2012; Ng'habi et al., 2018).

### **1.8.3 IVM interacts with ligand gated ion channels, but the cause of resistance is unknown**

The molecular basis of IVM resistance is likely to be multi-genic, and genes or mutations conferring resistance may vary between populations (Kotze et al., 2014). Although many studies have been conducted to date, there has been no consensus of a single dominant gene or single nucleotide polymorphism (SNP) conferring IVM resistance. In contrast, studies have found strong associations between non-synonymous SNPs or indels present within target genes and anthelmintic resistance for both BZ (Kwa et al., 1995) and LEV (Barrère et al., 2014; Blanchard et al., 2018). The model nematode, *Caenorhabditis elegans*, is used to study anthelmintic resistance within laboratory settings as it is easily manipulated and maintained, unlike ruminant parasitic nematodes. IVM was

shown to irreversibly activate a *C. elegans* glutamate gated chloride channel subunit, (GluCl), *glc-1* (*GluClA*) expressed in *Xenopus* oocytes (Cully et al., 1994), but a second GluCl tested (*glc-2*, *GluClB*) was insensitive. These channels have been identified in invertebrates but not in the mammalian host (Wolstenholme, 2012). Initially, these target genes received much attention, and it was shown using *C. elegans* that IVM potentiated the effect of glutamate, causing paralysis of worms. The action of IVM was correlated with the presence of bound IVM in membrane fractions, suggesting that the channels were expressed in the cell membranes of *C. elegans* (Arena et al., 1995). A further GluCl gene was identified, which was post-synaptically expressed on the muscles of the pharynx and some head and anal motor neurons, and which was necessary for the pharynx muscles to respond to glutamate (Dent et al., 1997). The authors found that mutations in this gene alone, *avr-15*, were not sufficient to induce IVM resistance, but that worms with mutations in this gene could be further selected with IVM to produce resistant individuals carrying mutations in both *avr-15* and a second unknown gene (Dent et al., 1997). They suggested that the main nematocidal property of IVM was its ability to paralyse the pharynx of worms, inducing starvation and preventing development to the next lifecycle stage. They observed that somatically paralysed adults of *C. elegans* were still viable, and hypothesised that the importance of somatic paralysis in a parasitic nematode would depend on its location in the host (Dent et al., 1997). Just three years later, the authors published information showing that, in *C. elegans*, IVM resistance increases 4000 fold following null mutations (variants which result in loss of function of the gene), in three GluCl genes; *glc-1*, *avr-14* and *avr-15* (Dent et al., 2000).

Early studies investigating the *avr-14* orthologue in the cattle parasitic nematode *C. oncophora*, identified three SNPs (E114G, V235A and L256F) (Njue and Prichard, 2004). Following expression of these IVM sensitive and IVM resistance-associated haplotypes in *Xenopus* oocytes, resistance-associated haplotypes were shown to be less sensitive to glutamate and to IVM. However, the fold changes were small, (2.6 fold increase in the EC<sub>50</sub>), and were predominantly related to the L256F SNP (Njue et al., 2004). Consistent with Dent et al. (2000), this would suggest that mutations in single genes confer little or no resistance. Both Cully et al. (1994) and Njue et al. (2004) found increased sensitivity to glutamate

when heteromeric channels (*glc-1* and *glc-2*; *avr-14* and *glc-2*) were expressed in *Xenopus* oocytes rather than homomeric channels, and this may partially underlie the requirement for multiple gene mutations. In 2011, a study using IVM resistant Belgian isolates of *C. oncophora* and *O. ostertagi* failed to detect the L256F polymorphism, but they did note that the *avr-14* gene was highly polymorphic, with 32 alleles in a 190 bp region identified in a sensitive isolate and reduced allele diversity in resistant isolates. Additionally, reduced expression of both *avr-14A* and *avr-14B* in IVM resistant adults of *C. oncophora* was observed (El-Abdellati et al., 2011). In contrast, a study by Martinez-Valladares et al. (2012b) searched for the presence of the L256F SNP within the *T. circumcincta* *avr-14* gene. They failed to find either significant evidence of selection by IVM, or the presence of the SNP, and concluded that it was unlikely that *avr-14* played a significant role in IVM resistance in *T. circumcincta*. One allele was significantly increased in frequency within an MTci5 adult population following IVM treatment compared to the IVM susceptible MTci1 population. Although non-synonymous SNPs were identified within *avr-14B*, all were at a low frequency (Martinez-Valladares et al., 2012b).

In the parasitic nematode *H. contortus*, GluCl genes include *avr-14*, *avr-15*, *glc-2*, *glc-3*, *glc-4*, *glc-5* and *glc-6* (McCavera et al., 2007; Glendinning et al., 2011). However, no orthologue of *glc-1* is present (Laing et al., 2013). A few studies have shown associations between changes in GluCl and IVM phenotype, however others have not identified differences. For example, Blackhall et al. (1998) selected two *H. contortus* IVM resistant strains, using sub-optimal dosing over 17 generations, from two parental isolates. The resistant strains survived approximately 10-fold higher IVM concentrations than that required to kill 95% of the parental isolates. These selected strains were compared to un-selected, sensitive strains derived from the same parental isolates. Following SSCP analysis, one GluCl allele was significantly increased in frequency in the resistant strain compared to the corresponding susceptible strain for one pair. However, no significant change was noted for the other strain (Blackhall et al., 1998). The gene was later identified as *Hco-glc-5* (Beech et al., 2010).

In a similar way, differences were noted in allele frequency in some candidate IVM resistance genes in a study comparing four UK farm populations of *H.*



*contortus* but these were not associated with IVM resistance (Laing et al., 2016). No significant differences were noted between UGA/2004, a resistant isolate, and US/HcS, a susceptible isolate (Williamson et al., 2011). Neither were *Hco-glc-5* haplotypes linked with IVM resistance in backcrosses between the inbred susceptible strain MHco3(ISE), and two resistant isolates, MHco4(WRS) and MHco10(CAVR) (Rezansoff et al., 2016). Such changes more likely reflect diversity at this locus between populations therefore, rather than IVM selection. However, a 0.59 fold reduction in mRNA expression of *Hco-glc-5* in UGA/2004 compared with the US/HcS isolate was determined to be significant (Williamson et al., 2011), but, again, could reflect differences between isolates rather than IVM selection. Using the triple null mutant, IVM resistant *C. elegans* strain DA1316 (Dent et al., 2000), Glendinning et al. (2011) were able to demonstrate rescue of IVM induced paralysis using transgenic expression of *Hco-avr-14b* and partial rescue with *Hco-glc-5*, indicating IVM sensitivity of these channel subunits. Nevertheless, neither Laing et al. (2016), nor Rezansoff et al. (2016), found associations between *Hco-avr-14b* haplotypes and IVM resistance phenotype. Collectively, although GluCl's are targets of IVM, studies to date do not suggest a major or consistent role for GluCl's in mediating IVM resistance in parasitic nematodes of ruminants either in laboratory isolates or in field populations.

Other possible candidate IVM resistance genes include those of the ATP-binding cassette transporter family (ABC transporters), of which the most frequently investigated in parasitic nematodes are the P-glycoproteins (P-gp). These are a family of multi-drug transporters which are involved in efflux of xenobiotics from cells (Lespine et al., 2012). Originally identified in multi-drug resistant cancer cells (Juliano and Ling, 1976), interaction of IVM with P-gps has been shown in both mammalian cell expression systems and in *C. elegans* (Godoy et al., 2015; Janssen et al., 2015; Godoy et al., 2016; Mani et al., 2016). Transgenic expression of parasitic nematode P-gps in mammalian cells, including *Dirofilaria immitis* PGP-11 (Mani et al., 2016) and *H. contortus* PGP-2 and PGP-9.1 (Godoy et al., 2015, 2016), demonstrated inhibition of fluorophore accumulation by IVM. *Parascaris pgp-11*, transgenically expressed in a *pgp-11* deficient *C. elegans*, was shown to confer a 3.2 - 4.6 fold increase in IVM resistance relative to the deficient strain in a thrashing assay (Janssen et al.,

2015). A thrashing assay is used in *C. elegans* to measure the effect of a drug or a mutation on the ability of the worms to move in liquid - the number of lateral movements of the body are counted and compared to controls. In *C. elegans*, 15 P-gps are present (Sheps et al., 2004). Selective culture of an IVM resistant *C. elegans* strain on gradually increasing IVM concentrations up to 10 ng/ml, resulted in increased expression of P-gps, especially *pgp-1*, and another multi-drug transporter, *mrp-1* (James and Davey, 2009). In contrast, rapid selection of a *H. contortus* resistant strain (parental isolate MHco3(ISE)), did not identify changes in P-gp expression (Williamson and Wolstenholme, 2012). Different P-gps may be important in anthelmintic resistance in different parasitic species, including *pgp-2* in *H. contortus* (Lloberas et al., 2013), *pgp-9* in *T. circumcincta* (Dicker et al., 2011b), and *pgp-11* in *Parascaris* (Janssen et al., 2015). Several studies have found associations between constitutive and induced mRNA expression levels of P-gp genes and IVM resistance phenotypes in *H. contortus* (Xu et al., 1998; Prichard and Roulet, 2007; Williamson et al., 2011; Lloberas et al., 2013). Time dependent differences in P-gp expression have been observed (Lloberas et al., 2013), and 14 days post-IVM no differences were noted in a resistant *H. contortus* strain between untreated, treated worms or worms treated at ten times the recommended dose (Alvarez et al., 2015).

In *T. circumcincta*, 11 partial P-gp sequences have been detected (Dicker et al., 2011b). Upregulated expression of *pgp-9* but downregulated expression of *pgp-2* was associated with the BZ, LEV and IVM resistant MTci5 isolate, compared with the sensitive isolate, MTci2 (Dicker et al., 2011b). Significantly reduced (0.12 to 0.17 fold) expression of *pgp-2* was identified in eggs, L1, ex-sheathed L3 and L4, but not in adults of MTci5 compared with MTci2 (Dicker et al., 2011b). In contrast, expression levels of *pgp-9* were significantly increased (5.06 to 55.27 fold) in all life stages of MTci5 compared with MTci2, most noticeably in eggs (Dicker et al., 2011b).

Both synonymous and non-synonymous SNPs have been found within the *Tci-pgp-9* gene, and changes in haplotype frequencies have been noted, although the significance of these in relation to resistance has yet to be conclusively shown (Choi et al., 2017; Turnbull et al., 2018). Comparing the full-length *pgp-9* cDNA (~3.8 kb) of MTci2 and MTci5, nine non-synonymous SNPs, corresponding to six

amino acid changes, were identified (Turnbull et al., 2018). A further four amino acid changes within the same region had been previously identified by Bisset (2007). Of these ten amino acid substitutions, only four were consistently identified with an IVM resistant phenotype (A662T, T663A, A664T, R697S) (Turnbull et al., 2018). Using individual worm gDNA qPCR, and pooled read counts, Turnbull et al. (2018), and Choi et al. (2017) found evidence for a potential increase in copy number of *Tci-pgp-9* over and above that expected for a single copy gene in a diploid, heterozygous individual. Overall, Turnbull et al. (2018) identified a high level of within-population diversity at the *Tci-pgp-9* locus. Up to ten, seven and five *pgp-9* haplotypes were identified in individuals of MTci2, MTci5 and post-IVM MTci5 populations, respectively (Turnbull et al., 2018). The authors concluded that these results could indicate selection at the *pgp-9* locus both by anthelmintics in general and IVM in particular.

Studies investigating the potency of IVM *in vitro* with *T. circumcincta* and *H. contortus* L1, found IVM to be more effective against both susceptible (MTci3 and MHco3(ISE)) and resistant (MTci4 and MHco4(WRS)) isolates, when worms were exposed to P-gp inhibitors (Bartley et al., 2009). Despite these findings, fold changes associated with P-gps are often small, with several studies unable to detect associations between P-gps and IVM resistance (Dicker et al., 2011a; Williamson and Wolstenholme, 2012; Alvarez et al., 2015; Rezansoff et al., 2016).

Other genes have also been investigated for their role in IVM resistance in nematodes. In the study by Dent et al. (2000), loss of wild type innexins (*unc-7* and *unc-9*, molecules forming gap junctions and involved in cell to cell transmission), was found to enhance IVM resistance in *C. elegans*. An amphidial, dye-filling defective gene (i.e. reduced uptake of Dil stain in amphids), *osm-1*, was also associated with resistance in *C. elegans* (Dent et al., 2000). In several laboratory *H. contortus* isolates a specific haplotype of *Hco-dyf-7*, another dye-filling defective gene, containing 15 SNPs, was found to be associated with an IVM resistance phenotype (Urdaneta-Marquez et al., 2014). Nevertheless, no association of *dyf-7* genotype and IVM selection was found in four *H. contortus* field populations within the UK (Laing et al., 2016). Recently, a study investigating the association between reduced Dil uptake by *C. elegans*

amphidial neurons and IVM resistance identified 17 mutants in a forward genetic screen which were able to survive on 10 nM IVM agar plates (Page, 2018). Of these, just two had variable filling of the amphids (the rest were dye-filling defective) and these were identified to contain mutations in dynein heavy chain genes, *che-3* and *dhc-3*. IVM resistance in the wild type sensitive strain, N2, was induced by RNAi of either gene (Page, 2018). By screening previously identified dye-filling defective mutant strains, Page (2018) was able to confirm resistance to 10 nM IVM in ten strains, including a strain defective for *dyf-7*, but not a strain defective for *osm-1*. Finally, other ligand gated ion channels (LGIC) may also be important in IVM resistance, as IVM is known to antagonise *Ascaris* GABA receptors in L1 muscle cells (Holden-Dye and Walker, 1990).

Since the advent of next generation sequencing (NGS) technologies, the opportunity to conduct genome wide association studies (GWAS) has become increasingly available as sequencing prices have reduced (Park and Kim, 2016). In 2017, the first GWAS in *T. circumcincta* investigating BZ, LEV and IVM resistance was published (Choi et al., 2017). A susceptible and a resistant NZ strain were used. The multi-drug (BZ, LEV and IVM) resistant strain had been isolated from the field and then subsequently selected in the lab for five generations, in an attempt to reduce diversity and increase homozygosity of resistance alleles. Backcrossing of these strains with *in vivo* anthelmintic selection had been carried out and both whole genome re-sequencing (WGS) and double-digest restriction-site-associated DNA sequencing (ddRAD-Seq) was performed (Choi et al., 2017). A separate study, using 2b-RAD, investigated IVM resistance in two laboratory isolates of *H. contortus* (Luo et al., 2017). Recently, WGS of backcrosses of two resistant *H. contortus* isolates to a susceptible isolate, with IVM selection at each of four generations has also been performed (Doyle et al., 2019). Importantly, none of these studies have found associations between any of the previously studied ‘candidate’ genes and IVM resistance, with the exception of *pgp-9* in *T. circumcincta*. Here, there was a weak correlation between *Tci-pgp-9* and IVM resistance, with increased copy number and distinct haplotypes detected within the multi-drug resistant population compared to the susceptible inbred strain (Choi et al., 2017). Instead, various genes, including a LGIC, *Tci-lgc-54* (Choi et al., 2017) and *Tci-mrp-6* (an ABC

transporter) (Choi et al., 2017) were implicated in IVM resistance in *T. circumcincta*.

It is important to note that most studies have investigated IVM resistance using multi-drug resistant isolates. For example, MTci5 is resistant to BZ, LEV and IVM. These factors may impact on the outcome of a study, as mechanisms identified may be equally beneficial to the nematode in resisting the effect of multiple anthelmintic classes. This is especially true of the P-gps, recognised as a method of multi-drug resistance in cancer studies (Lehne, 2000). In addition, many studies compare isolates and parasite populations obtained from discrete geographical regions or hosts, both of which will impact on genetic selection pressures, leading to differences in genotype which may have little bearing on IVM sensitivity (Sallé et al., 2019). Lastly, nematode parasite populations are highly diverse and sub-populations of drug resistant worms may exist within the overall multi-drug resistant population (Sarai et al., 2014). This phenomenon may confound results if not accounted for, however it can also be exploited to detect regions of the genome specifically under selection by IVM (Doyle et al., 2019).

## 1.9 Benzimidazole Resistance

At the time of writing, molecular detection of polymorphisms is only established for BZ resistance, with the understanding that non-synonymous SNPs in the  $\beta$ -tubulin isotype-1 gene are capable of conferring resistance in parasitic strongyles (Coles et al., 2006; Beech et al., 2011). These include SNPs in codons F167Y (tTc > tAc) (Silvestre and Cabaret, 2002), E198A/L (gAa > gCa or GAa/GAG > TTa) (Ghisi et al., 2007; Rufener et al., 2009a; Redman et al., 2015) and, most commonly, in codon F200Y (tTc > tAc) (Kwa et al., 1995). Computer modelling has suggested that the presence of tyrosine at codon 200 prevents BZ from binding  $\beta$ -tubulin isotype-1, and is the amino acid found at this position in mammalian  $\beta$ -tubulin isotype-1 (Robinson et al., 2004).

None of the molecular methods commonly in use allow interpretation of all three codons at once, necessitating at least two PCRs. In addition, some methods are more limited in their interpretation than others. Pyrosequencing conditions for both *H. contortus* (von Samson-Himmelstjerna et al., 2009b) and

*T. circumcincta* (Skuce et al., 2010) have been published, in which all SNPs are sequenced in two separate reactions; these and similar assays have been employed in several field studies (von Samson-Himmelstjerna et al., 2009b; Skuce et al., 2010; Martinez-Valladares et al., 2012a; Ramünke et al., 2016). In contrast, allele-specific PCRs have been reported for codon 200 (Elard and Humbert, 1999; Silvestre and Humbert, 2000), and for alanine at codon 198 (Rufener et al., 2009a) but not for codon 167, requiring separate PCRs and limiting interpretation of results. Furthermore the allele-specific PCRs are fickle and require highly stringent lab techniques, with accurate primer concentrations, and fresh L3 (Coles et al., 2006). As for the *in vivo* and *in vitro* tests, molecular tests on pools of larvae can be confounded by species identity. One particular allele-specific PCR, including a speciation step in the protocol, was described by Silvestre and Humbert (2000). Species identity was determined using a restriction fragment length polymorphism PCR, prior to detection of the codon 200 SNP allele using  $\beta$ -tubulin isotype-1 species-specific primers. However, this can be difficult to interpret due to the presence of multiple bands in the speciation PCR. NGS is now being used to swiftly diagnose and investigate BZ resistance in pools of mixed species individuals (Avramenko et al., 2019), which can be comparatively easy to set-up, but expensive to perform. A potential inexpensive molecular test for field use may be the Loop-mediated isothermal amplification (LAMP) assay which has the ability to detect three alleles of a SNP that confers resistance to the BZ fungicide carbendazim (Duan et al., 2018). These fungal LAMP assays appear to be both sensitive and specific, even in the presence of large amounts of non-organismal DNA (Duan et al., 2016). It is likely that other BZ resistance mechanisms exist (Sangster et al., 1998; Beech et al., 2011; Kotze et al., 2014; Hahnel et al., 2018), but detection of these has been limited.

## 1.10 Levamisole Resistance

LEV induces a spastic paralysis in nematodes by binding to acetylcholine receptors (AChR) on nematode somatic muscle cells (Lewis et al., 1980). There are many types of AChR, which vary between species (Qian et al., 2008; Neveu et al., 2010; Boulin et al., 2011) and there is inherent variability in their ligand sensitivity and specificity (Qian et al., 2006). There also appears to be a degree of plasticity or redundancy within the nAChR subunit gene family, which may

complicate the understanding of LEV resistance. In *C. elegans*, the LEV sensitive AChR is pentameric, composed of three  $\alpha$ -subunits: LEV-8, UNC-38 and UNC-63 (Culetto et al., 2004; Towers et al., 2005), and two non- $\alpha$  subunits, LEV-1 and UNC-29 (Fleming et al., 1997). Three ancillary factors, RIC-3, UNC-50 and UNC-74 are required for functional expression in *Xenopus* oocytes (Boulin et al., 2008). However, in the parasitic nematodes, *H. contortus* and *T. circumcincta*, no *lev-8* orthologue has been identified (Neveu et al., 2010; Blanchard et al., 2018). A functional LEV sensitive receptor was produced in *Xenopus* oocytes using *Hco-unc-38*, *Hco-unc-63*, *Hco-unc-29*, and *Hco-acr-8* (Boulin et al., 2011). Additional expression of *Hco-lev-1* made little difference to the current detected following application of LEV to the cells. The absence of *Hco-acr-8* greatly reduced the current response, but did not eliminate it (Boulin et al., 2011). Homologues of these genes have been identified in *T. circumcincta*, although functional expression has not yet been performed (Walker et al., 2001; Neveu et al., 2010; Boulin et al., 2011). Transgenically expressed *Hco-unc-38* cDNA, was sufficient to fully rescue the LEV resistant phenotype (with associated behavioural abnormalities) of a *C. elegans* null mutant, but *Hco-unc-29* was only able to partially rescue a *Cel-unc-29* null mutant (Sloan et al., 2015). However, the authors used only one isotype of *Hco-unc-29*, while four isoforms of *unc-29* are recognised in both *H. contortus* and *T. circumcincta* (Neveu et al., 2010). Finally, Blanchard et al. (2018) produced evidence that the transgenic expression of *Hco-acr-8* in body wall muscle was able to restore LEV sensitivity in a *C. elegans* mutant strain lacking *Cel-lev-8* and resistant to LEV. This indicates that *Cel-acr-8* may not naturally form functional LEV sensitive receptors in *C. elegans* in the absence of *Cel-lev-8*, despite the formation of such receptors in *Xenopus* oocytes. Or, if it does, that it is less responsive to LEV than when expressed in *Xenopus* oocytes (Blanchard et al., 2018). Finally, knock-down of *acr-8* using siRNA in *H. contortus* L2 induced a LEV resistance phenotype compared to control larvae (Blanchard et al., 2018).

LEV resistance in parasitic nematodes has been predominantly associated with truncated forms of alpha-subunits of AChRs. A truncated isoform of *acr-8*, termed *acr-8b* in *H. contortus* (Fauvin et al., 2010) and also a truncated isoform of *unc-63*, termed *unc-63b* in *H. contortus*, *T. circumcincta* and *Trichostrongylus colubriformis* (Neveu et al., 2010) have been identified. Again using *Xenopus*

oocytes as an expression system, the presence of *Hco-unc-63* was shown to be required for a functional receptor, however the co-expression of *Hco-unc-63b* was found to have a dominant negative effect, greatly reducing the response of the receptor (Boulin et al., 2011). The presence of an indel within the second intron of *Hco-acr-8* has since been associated with the *Hco-acr-8b* truncated transcript in various isolates, recognising that the transcript is formed from exons 1 and 2 of *Hco-acr-8*, with an additional third exon formed from sequence found within intron 2 (Barrère et al., 2014). Nonetheless, resistance appears complex, with both LEV-resistant and LEV-susceptible isolates containing the deletion (Barrère et al., 2014). Also, expression patterns of the various subunits, including *Hco-acr-8b* and *Hco-unc-63b*, and the three ancillary factor genes vary between lifecycle stages within an isolate and between isolates (Sarai et al., 2013, 2014). In addition, the relative expression of these subunits is not conserved in three separate *H. contortus* multi-drug resistant isolates compared to a susceptible isolate (Sarai et al., 2013). Some of this variation may be explained by sub-populations of worms with different LEV phenotype within an overall population. Higher level *in vitro* resistance of *H. contortus* Wallangra L3 was associated with more consistently reduced expression of AChR subunits and ancillary genes compared to the whole population (Sarai et al., 2014). Various amino acid changes have been identified in AChR subunits in *H. contortus*, although prevalence between three resistant isolates varied (Sarai et al., 2013). In *T. circumcincta*, differences in frequency of *Tci-unc-38* (*tca-1*) alleles between a resistant isolate and two susceptible isolates were observed, including amino acid changes, but no association of any particular allele with LEV resistance was identified (Walker et al., 2001). Interestingly, in *C. elegans*, a single basepair change in the first intron of *Cel-lev-8* corresponds to the presence of two alternatively spliced isoforms, with premature stop codons (Towers et al., 2005). *lev-8* mutant worms are partially resistant to LEV, and can be partially rescued to wildtype sensitivity by extrachromosomal expression of *Cel-lev-8* (Towers et al., 2005). LEV resistance is therefore potentially complex, with at least 11 genes originally identified in early *C. elegans* studies associated with variation in resistance phenotype between mutants (Fleming et al., 1997). Furthermore, a considerable number of genes are involved in the pathway enabling functional responses to ACh and LEV (Martin et al., 2012). Although there is good evidence for the correlation of indels in *acr-8* with LEV sensitivity

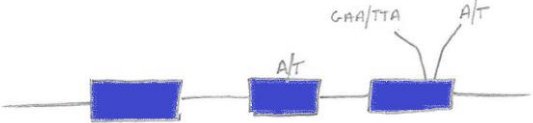

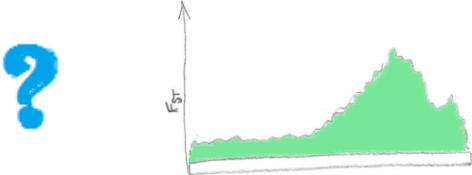
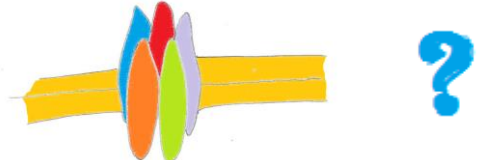
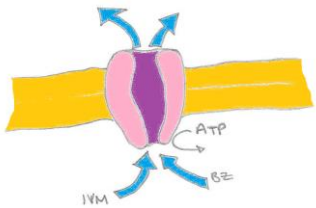


in *H. contortus* (Fauvin et al., 2010; Blanchard et al., 2018), the deletion appears to vary between isolates and has not been identified in other strongyle species.

## 1.11 Monepantel Resistance

Before its release onto the market, Kaminsky et al. (2008) determined the genes associated with MPTL resistance in *C. elegans* (*acr-23*) and *H. contortus* (*des-2* and *acr-23*). Subsequently, *acr-23* was renamed *Hco-mptl-1* (Rufener et al., 2009b). MPTL resistant *T. circumcincta* strains have also been generated at the MRI: MTci2-11, MTci5-13 and MTci7-12, with sub-therapeutic dosing (Bartley et al., 2015). Of concern was the observation that the MPTL resistant *T. circumcincta* strain, MTci7-12 appeared to establish patency more rapidly and to be more fecund, than the parental strain (MTci7) from which it was derived (Bartley et al., 2015). In another study, rapid selection of a resistant *H. contortus* strain was possible in the laboratory after isolation of a field isolate obtained following a FECRT, which showed an efficacy of 99.2%, from which offspring of the survivors were used to infect a single sheep (Raza et al., 2016c). A FECRT was performed and MPTL efficacy was found to be only 15% effective (Raza et al., 2016c). Interestingly, LDTs demonstrated the presence of at least two separate sub-populations. This may indicate two or more mechanisms of resistance, conferring low and high level (more than 1000-fold increase of the ED<sub>50</sub> compared to a susceptible isolate) resistance to MPTL *in vitro* (Raza et al., 2016c). Sequencing of the resistance-associated gene, *Tci-mptl-1*, in both lab and field populations, has confirmed positive selection at this locus, noting a reduction in allele frequency from many haplotypes pre-treatment, to just one major haplotype post-treatment (Turnbull et al., 2019). However the causative polymorphism is unknown, and the alleles under selection differed between isolates, suggesting multiple genetic backgrounds are associated with resistance. Work is still ongoing, but it may be that a diagnostic test of MPTL resistance will need to encompass more than one variant. Recently a GWAS using a *H. contortus* genetic cross confirmed a single region surrounding *Hco-mptl-1* is under selection by MPTL (Niciura et al., 2019).

Table 1.4: Mechanisms of resistance in parasitic strongyles.

Anthelmintic	Mechanism of resistance in parasitic strongyles		Potential mechanisms of resistance
BZ		<p><math>\beta</math>-tubulin isotype-1 SNPs; F167Y, E198A/L, F200Y (Kwa et al., 1995; Silvestre and Cabaret, 2002; Rufener et al., 2009a)</p>	<p><math>\beta</math>-tubulin isotype-2 (Kwa et al., 1993) Other loci (Hahnel et al., 2018)</p>
LEV		<p>AChR truncated proteins; <i>acr-8/acr-8b</i>, <i>unc-63/unc-63b</i> (Fauvin et al., 2010; Neveu et al., 2010; Boulin et al., 2011; Barrère et al., 2014)</p>	<p>Other AChRs or genes involved in the LEV sensitive AChR pathway (Fleming et al., 1997; Walker et al., 2001; Towers et al., 2005; Martin et al., 2012)</p>
IVM		<p>Unknown Gene(s) within a region on Chromosome V in <i>H. contortus</i> (Doyle et al., 2019)</p>	<p>GluCl<sub>s</sub> (Dent et al., 2000) Amphidial genes (Dent et al., 2000; Urdaneta-Marquez et al., 2014; Page, 2018) Innexins (Dent et al., 2000)</p>
MON		<p>AChRs, mechanism unknown; <i>mptl-1/acr-23</i>, <i>des-2</i> (Kaminsky et al., 2008; Rufener et al., 2009b)</p>	<p>AChR <i>deg-3</i> (Rufener et al., 2009b)</p>
Multi-drug resistance		<p>P-glycoproteins (Blackhall et al., 2008; Bartley et al., 2009; Dicker et al., 2011b; Raza et al., 2016b; Turnbull et al., 2018)</p>	<p>Tolerance; transcription factors (Ménez et al., 2019), lipid metabolism (Laing et al., 2012), drug metabolism (Fontaine and Choe, 2018; Stuchlikova et al., 2018). Amphids (Page, 2018)</p>

## 1.12 What factors affect the genetic footprint of resistance?

Anthelmintic resistance can arise by one of two methods. Either resistance conferring mutations are already present within a population and these are selected for by treatment, or mutations arise *de novo* during active anthelmintic use on a farm. A study by Redman et al. (2015), sequencing a region of the  $\beta$ -tubulin isotype-1 gene, noted that mutations arose independently on several farms, as had also been found in another study of BZ resistance (Skuce et al., 2010). In addition, Redman et al. (2015) demonstrated the spread of BZ resistance between several UK farms, likely due to sheep movement, which was particularly obvious for the less common mutation noted at codon E198L. Strongyles are sexually reproducing, and *H. contortus* has been shown to be polyandrous, with occasional fully or partially triploid individuals (Doyle et al., 2018). Mutation rates are expected to be high, despite comparatively low recombination rates (Doyle et al., 2018). These traits are reflected in the relatively high nucleotide diversity ( $\pi$ ) estimates of 0.022 to 0.038 identified for autosomes of *H. contortus* isolates (Doyle et al., 2019). These stand in contrast to nucleotide diversity estimates for the hermaphrodite *C. elegans*, which are much lower and were calculated to be only 0.00068 for one sample population (Cutter, 2006), despite a recombination rate almost double that of *H. contortus* (Barnes et al., 1995). This high nucleotide diversity, the probability that two sequences will differ at a given genomic site, provides the potential for dilution of resistance SNPs within the population as a whole, in the absence of drug selection as new mutations occur. However, it also increases the chance of a resistance-conferring SNP arising *de novo*. In the presence of drug, positive selection occurs, and the number of alleles within the population decreases (Lubega et al., 1994; Redman et al., 2012). Dependent on the number of independent origins of resistance, the number of alleles retained will vary. If there is a hard sweep, just one or two genetic backgrounds carrying the resistant allele will be selected (Barnes et al., 2017). In contrast, a soft sweep will occur when many genetic backgrounds contain the resistant allele (Redman et al., 2015). However, with a high mutation rate, and a high recombination rate, the time since the variant arose will greatly affect how easy it is to detect such a selection pressure. For example, ancient mutations are likely to now be present

on many different genetic backgrounds and a hard sweep could resemble a soft sweep. As the rate of recombination varies between different chromosomes of *H. contortus* (Doyle et al., 2018), the position of a variant within the genome is also important as to how rapidly the genetic background surrounding the variant will diversify. Recombination rate also varies along individual chromosomes, in both *H. contortus* (Doyle et al., 2018) and *C. elegans* (Rockman and Kruglyak, 2009), such that the position within a chromosome may also effect the size of the genetic footprint identified. Whether a resistance conferring mutation is dominant or recessive will also affect the selection pressure seen (Smith and Haigh, 1974). Finally, the potential for dilution of resistance alleles will be impacted upon by the size of the population unexposed to drug, known as the population *in refugia* (Martin et al., 1981; van Wyk, 2001).

### **1.13 Why might anthelmintic resistance have arisen so rapidly?**

Nematode parasites of sheep are generally highly fecund, with each *T. circumcincta* adult able to lay about 100-200 eggs per day (Cole, 1986), and each *H. contortus* female capable of producing approximately 5000 eggs per day (Coyne and Smith, 1992). Large parasite population sizes, and the delay between treatment of a host with an anthelmintic and patency of new infection, could rapidly increase the prevalence of resistance alleles on pasture. In the past, farmers were advised to ‘treat and move’ their sheep onto clean pasture, a management practice which likely exacerbated development of resistance. All farmers are currently advised to quarantine and treat new stock on pasture not routinely grazed by other animals on the farm, or indoors (especially for the first 48 hours) (Abbott et al., 2012). However, there is no guarantee that this will be performed (Jack et al., 2017), nor that multi-drug resistant parasites are not present, allowing resistant parasites to be introduced to the flock. The practice of under-dosing sheep, by inaccurate estimation of an animal’s weight, poor calibration of a dosing gun, or following inefficient dosing such that an animal swallows only part of the required dose, is also thought to be a major factor in the development and spread of resistance (McMahon et al., 2013c). It has been proposed that the use of sub-optimal doses allows for more rapid development of resistance, but by the selection of genetic mutations which may not confer full resistance against the correct dose. For example, in *C. elegans*, generational

selection of worms using increasing doses of IVM identified selection of P-gps (James and Davey, 2009), in contrast to GluCl<sub>s</sub>. Alternatively, sub-optimal dosing may allow parasites heterozygous for a recessive resistance-conferring mutation, to survive and contribute towards the next generation of parasites. In the UK, where we have a stratified sheep industry, the spread of resistance alleles between farms appears to be inevitable.

## 1.14 What are the concerns with anthelmintic resistance?

PGE raises both welfare and economic concerns. For UK farmers, as for many others, the inability to control PGE calls into question the sustainability of sheep farming (Lovatt and Stubbings, 2018). Some farms have already closed due to a lack of control options in the face of multiple anthelmintic resistance (Sargison et al., 2005). Following BREXIT, it is likely that many farmers will fail to make a profit unless they are able to make considerable changes to their farm management and production costs/outgoings (AHDB, 2017). Although other production diseases obviously impact flocks, PGE is ubiquitous and most farmers rely heavily on anthelmintics to control parasites and maintain reasonable weight gains on farm (Burgess et al., 2012). Whilst in the UK the cost of rising anthelmintic resistance will predominantly be felt by individual farmers, their families and their flocks, on a worldwide scale the sustainability of food production is a key concern (Fitzpatrick, 2013; Morgan et al., 2013). Many subsistence farmers rely on small ruminants to provide both milk and meat (Alary et al., 2011) and PGE has been shown to considerably affect both these production traits (Mavrot et al., 2015). Anthelmintic resistance in these conditions contributes to human malnutrition and subsequently a rise in endemic diseases of both livestock and humans.

As anthelmintic resistance increases within strongyle populations on a holding, the benefit of using these products reduces until clinically apparent anthelmintic resistance is seen (Sargison, 2011). In addition, although for the BZ, LEV and macrocyclic lactones, no apparent benefit has been noted for anthelmintic resistant parasites, there are also no reports of a fitness cost of anthelmintic resistance (Elard et al., 1998; Leathwick, 2013). In contrast, one MPTL resistant *T. circumcincta* laboratory strain did appear to have enhanced parasite fitness traits compared to its MPTL sensitive parental population (Bartley et al., 2015).

These differences however were not seen in two other selected strains and may be related to genetic drift, or genetic hitch-hiking of fitness traits which are isolate specific, rather than a change directly related to anthelmintic resistance (Maingi et al., 1990). Anthelmintic resistance is not only present in sheep nematodes but is also a serious problem in: *Fasciola hepatica* populations (Kelley et al., 2016), cattle nematodes (*O. ostertagi*, *C. oncophora*) (Kaplan and Vidyashankar, 2012), human nematodes (sub-optimal responses noted to IVM in *Onchocerca volvulus*) (Awadzi et al., 2004), canine heartworm (*D. immitis*) (Bourguinat et al., 2015), equine strongyles (Nielsen et al., 2014), and *P. ovis*, where macrocyclic lactones are used (Sturgess-Osborne et al., 2019). Overall this trend highlights the capability of these parasites to survive and adapt to changing circumstances. Although livestock farming is still possible in the face of anthelmintic resistance, productivity will fall and, for many, this may make commercial farming and production of food for the human population untenable.

### **1.15 Is there hope in the face of anthelmintic resistance?**

Although no studies have yet shown that anthelmintic resistance associated genotypes may be lost from the population, several have demonstrated improvement in anthelmintic efficacy within a flock over time (Kenyon et al., 2013; Leathwick et al., 2015). This has followed use of newer and combination products over the use of older, single class anthelmintics (Leathwick et al., 2015), and the use of targeted selective treatment within a flock (Kenyon et al., 2013). In the study of Leathwick et al. (2015), no improvement in drug efficacy was observed for BZ anthelmintics, perhaps relating to fixation of BZ resistance within the population. Potentially, improvements for other anthelmintics might have arisen from a decrease in the prevalence of anthelmintic resistant genotypes within the parasite population as a whole, or from reduced proportions of these genotypes contributing to resistant phenotypes. For example, dilution by a susceptible *in refugia* population could lead to the presence of heterozygotes. Alternatively, improvement may simply be due to increased within host strongyle diversity with sensitive species (Melville et al., 2016), without any change in the prevalence of the resistance alleles within the resistant species population. If anthelmintic resistance is conferred by recessive alleles then the presence of a sensitive *in refugia* population may help to mitigate selection pressure by anthelmintic treatments. However, it will need to

be carefully managed as, without a fitness cost, any anthelmintic use will prioritise the survival of resistant genotypes over sensitive genotypes (van Wyk, 2001). This is true for either dominant or recessive genotypes. The former will confer a resistance phenotype in both heterozygotes and homozygotes. In contrast, recessive genotypes, unless there is sufficient refugia, will rapidly increase in the population to fixation as only homozygotes survive treatment.

However, anthelmintic resistance is not, by itself, a disease but merely a constraint to one management option of PGE (Sargison, 2014). With other management practices, such as co-grazing or rotational grazing with cattle (Jordan et al., 1988), or the judicious use of alternative forages and arable crops (Marley et al., 2003), it may be possible to better control larval numbers and strongyle diversity reducing the severity of PGE (Learmount et al., 2016b). Selective breeding of sheep enables selection for those resilient (high egg output, mild clinical signs) or resistant (low egg output), to *T. circumcincta*, by the use of recording on farm and through sire/maternal offspring monitoring (Morris et al., 2010). One such method, which may increase in use for this purpose, is that of monitoring the levels of lamb salivary IgA, with affinity for *T. circumcincta* L3 antigen *in vitro* (Fairlie-Clarke et al., 2019). As lambs are the main drivers of the multiplication of pasture larvae during the summer months, the ability to rapidly finish lambs earlier in the year is another way to better mitigate the effects of PGE. AHDB Beef and Lamb estimate that it is possible to finish lambs within 12 weeks, however in order for this to occur excellent pasture management is required to ensure the availability of sufficient, high quality feed (Laws and Genever, 2015). Some farmers choose to bring lambing forward, ensuring that lambs are finished before the warmer weather arrives. This avoids grazing young, susceptible lambs, at a time when eggs laid will rapidly develop to infective L3. Alternatively, it is possible to extend finishing time considerably, expecting lambs to remain on the farm far longer than normal, within a low input, slow output type system (van Wyk et al., 2003). Although this may not alleviate the effects of PGE, it can help to mitigate the production costs.

The use of vaccines (such as Barbervax<sup>®</sup> available for *H. contortus* <http://barbervax.com.au/>) and any future novel anthelmintics may assist with

the management of PGE (Smith et al., 2001). Although the *H. contortus* vaccine can greatly reduce reliance on anthelmintics, multiple doses are required over a breeding season (six for lambs and three to five for ewes) and some worms can survive the vaccine (Sallé et al., 2018). It is thus possible that any vaccine developed will still have a limited lifespan and/or will always require to be used alongside additional management tools. Furthermore, all vaccines to date are species specific, requiring management of other strongyle species contributing to PGE. The use of combination anthelmintic products may assist in reducing clinical disease, however the presence of dual-resistant parasites within a population limits the effectiveness of this strategy (Sargison et al., 2001). The proportion of strongyles within the overall population carrying resistance genotypes will ultimately depend on a vast multitude of factors. These include: the initial presence of the resistance genotypes within the population (Silvestre et al., 2000), the accuracy and frequency of dosing (Leathwick et al., 2015), the presence and size of the *in refugia* population (Martin et al., 1981), management factors (Sargison et al., 2007), host immune responses (Ahmed et al., 2015) and climatic conditions (Kenyon et al., 2009b; Rose et al., 2016) which may all impact parasite survival.

## **1.16 Investigating the genetic basis of drug resistance using next generation sequencing**

In 1977, Sanger sequencing was developed wherein the addition of complementary nucleotide bases (dNTPs) occurred until the addition of a ddNTP prevented further elongation of the strand. Identification of sequence lengths generated for each ddNTP base reaction allowed identification of the order of bases within a DNA sequence (Sanger et al., 1977). In 1990 this process was greatly improved through the use of capillary electrophoresis to separate and determine the size of the sequenced strands (Swerdlow and Gesteland, 1990). Although a significant advancement for understanding genetic mechanisms and allowing sequencing of genomes, the Sanger sequencing technique was slow and expensive, costing about \$300 million to sequence the human genome over several years (Bentley et al., 2008). Subsequently, NGS, also referred to as second generation sequencing, was developed. A range of NGS platforms are available, of which the Illumina is perhaps the most widely adopted (Bentley et al., 2008). These ‘massively parallel’ technologies have greatly increased the



potential for sequencing entire genomes and the cost per megabase sequenced has considerably reduced over the last decade, from approximately \$10,000 in 2001, to \$10 in 2008, and by 2017, human DNA could be sequenced for just \$0.01 per megabase (Wetterstrand, 2019). In addition, the Illumina sequencing method was considerably quicker than first generation sequencing methods, taking just eight weeks to sequence a human genome to a depth of 40x coverage (Bentley et al., 2008).

GWAS allow investigation into the genotype conferring a particular phenotype, without specifically targeting a single region or gene family. Instead, using an unbiased approach, GWAS look for regions of selection (or association) across the entire genome. These studies provide the means by which to identify novel regions under selection, without the constraint of pre-supposed theories as to the genetic basis of a trait. GWAS can be conducted using relatively few markers (SNPs or microsatellites, for example) spread across the genome. Alternatively, using whole genome sequencing techniques, the entire genome can be used to identify smaller regions of differentiation between two sample populations.

GWAS have successfully identified regions and genes associated with drug resistance in trematodes (e.g. *Schistosoma mansoni* and oxamniquine resistance, Valentim et al. (2013)), nematodes (e.g. *H. contortus* and IVM resistance, Doyle et al. (2019)), protozoa (e.g. *Plasmodium falciparum* and artemisinin resistance, Cheeseman et al. (2012)) and insects (e.g. *Anopheles funestus* and pyrethroid resistance, Barnes et al. (2017)). While some studies have identified novel regions under selection by a drug (Cheeseman et al., 2012; Takala-Harrison et al., 2013; Valentim et al., 2013; Doyle et al., 2019), others have aimed to confirm mechanisms already identified as the cause of drug resistance (Niciura et al., 2019), reducing the probability of the existence of additionally selected regions.

Several studies have been performed on parasites where clonal expansion within a host occurs. Clonal expansion has the major advantage of enabling a large number of identical parasites to be produced from a single resistant or susceptible individual as was recently demonstrated using *F. hepatica* (Hodgkinson et al., 2018). This provides a sufficient quantity of DNA for NGS from, essentially, a single individual, avoiding the use of error-inducing DNA

amplification methods. In contrast, mixed populations contain inherent diversity, which can complicate investigation of drug resistance. This is true both for sexually reproducing species, such as *H. contortus* (Sallé et al., 2019) and *T. circumcincta* (Choi et al., 2017), and for those capable of clonal expansion in addition to sexual reproduction, such as *F. hepatica* (Beesley et al., 2017). Although lab isolates of sexually reproducing parasites can be inbred, as has been performed with both *H. contortus* (Sargison et al., 2017) and *T. circumcincta* (Choi et al., 2017), diversity can still remain - often to a significant extent. In addition, resistance conferring variants identified in lab isolates may not reflect field population mutations or mechanisms. An ongoing study investigating triclabendazole resistance in *F. hepatica* has produced clonal populations of both resistant and susceptible field individuals (Hodgkinson et al., 2018). These have been used to generate field-derived genetic crosses to study triclabendazole resistance (Hodgkinson et al., 2013). This has provided both reduced diversity in addition to introgression and selection of resistance associated mutations, enhanced by knowledge of differences in parental SNPs between the resistant and susceptible clones. Others have used a genetic cross between susceptible and resistant lab isolates of the sexually reproducing *H. contortus* (Redman et al., 2012), or *T. circumcincta* (Choi et al., 2017) to introgress regions associated with anthelmintic resistance into a sensitive genetic background. However, Doyle et al. (2017) successfully used field populations of *Onchocerca volvulus*, without crossing or inbreeding of parasites, to identify novel regions of genetic differentiation based on IVM phenotype.

Often, additional testing using more outbred populations, or finer mapping with SNPs or microsatellites is employed to refine a region identified during a GWAS. One such example of this was by Valentim et al. (2013) investigating oxamniquine resistance in the trematode *S. mansoni*. A genetic cross was performed to identify a region under selection containing 184 genes. Following WGS to identify SNPs between the resistant and susceptible parental isolates within this region, a subset of these SNPs and additional microsatellites were utilised to refine the region. Having a list of just 16 potential candidates, they employed further genome-wide NGS (RNA-Seq) and additional experiments to identify the causative gene of oxamniquine resistance (Valentim et al., 2013). Using the knowledge that oxamniquine must be metabolised to its active form to

kill *S. mansoni*, they identified suitable genes by predicted function. Production of recombinant proteins and the use of a biochemical complementation assay confirmed a single sulfotransferase as the most likely gene candidate conferring oxamniquine resistance (Valentim et al., 2013). Subsequent studies used this information to further investigate oxamniquine resistance in the field (Chevalier et al., 2016; Chevalier et al., 2019). Hence, for the recessive oxamniquine resistance trait a GWAS using a genetic cross between clonal sample populations enabled identification of the region containing the gene responsible. However, identification of the gene itself relied on additional knowledge of how parasites mediate resistance to oxamniquine, which had been previously identified (Pica-Mattoccia et al., 2006).

There is a current need to identify markers of anthelmintic resistance in parasites for the macrocyclic lactones, in particular IVM, which is widely used as a broad-spectrum endectocide in both animals and humans (Campbell, 1985, 1991). Research investigating single genes has failed to find consistent evidence of a genetic marker within or between species (Kotze et al., 2014), and the mechanism of IVM resistance is unknown. Many studies are confounded by the use of isolates or strains originating from different geographic regions, or by comparing a susceptible isolate with a multi-drug resistant isolate (Laing et al., 2016).

Recent studies have used surgical techniques in sheep to generate genetic crosses between susceptible and resistant laboratory isolates of *H. contortus*, introgressing resistance markers into a sensitive genetic background. An initial study backcrossed MHco4(WRS) and MHco10(CAVR), two drug-resistant isolates from differing geographical origins, to MHco3(ISE), a susceptible isolate used in the genome project (Redman et al., 2012). The study identified a microsatellite marker, Hc\_8a20, which appeared to be linked to IVM resistance, and which has been identified in subsequent studies (Laing et al., 2016; Rezansoff et al., 2016). Recently, WGS of parental isolates and offspring of these *H. contortus* genetic crosses to investigate IVM resistance identified selection over a large region of Chromosome V (Doyle et al., 2019), which contained the Hc\_8a20 microsatellite. A third *H. contortus* genetic cross, using MHco3(ISE) crossed with MHco18(UGA), performed as part of work of the BUG Consortium

(<https://bugconsortium.wordpress.com/>) identified the same region on Chromosome V (Laing, *pers comm*).

Only a single GWAS investigating anthelmintic resistance in *T. circumcincta* has been performed. This employed NGS and backcrossed a triple resistant strain into a sensitive inbred strain (Choi et al., 2017). The study identified the  $\beta$ -tubulin isotype-1 locus (likely due to BZ selection), *Tci-pgp-9* and various LGICs in addition to other genes potentially involved in LEV or IVM resistance. However, the genome used (PRJNA72569, (Choi et al., 2017)), was highly fragmented and whether a few or many loci were under selection was not clear. In addition, it was not possible to determine genes solely related to IVM selection (Choi et al., 2017). Importantly, none of these studies have found candidate genes to be strongly associated with IVM resistant phenotypes. In contrast, introgression of MPTL resistance into a sensitive background confirmed the region around *Hco-mptl-1* to be under selection (Niciura et al., 2019).

Such GWAS have excellent potential and can clearly identify regions under selection specific to IVM through comparison with parental isolates or untreated progeny of the genetic cross. However, they can suffer two main limitations. Firstly, it is unknown whether the genetic basis of resistance is the same in field populations as in lab isolates. Also, generational selection of helminths in the lab takes considerable time, restricting the number of generations included in these studies. Thus the genetic footprints of selection can be quite large, containing many potentially important genes, such as was found with IVM resistance in *H. contortus* (Doyle et al., 2019).

To date many studies have investigated IVM resistance in both lab and field populations of sheep parasitic strongyles. However, despite rising prevalence of IVM resistance and decreasing efficacy of IVM and other anthelmintics on individual farms, no conclusion has been drawn as to the genetic basis of IVM resistance. For this reason studying how resistant genotypes increase in prevalence within a field population, and whether management practices are able to constrain development of resistance, is difficult. The need to identify genetic markers of IVM resistance in outbred field populations of *T. circumcincta* is the focus of this PhD. Such markers could potentially be used as diagnostic tests on farm, or within research studies to monitor IVM resistance in the field.

Initially a farm was closely monitored over a year to assess factors affecting detection of anthelmintic resistance using currently available methods (Farm 1, chapter 3). Following the identification of two appropriately resistant field populations (Farms 2 and 3), FECRTs were used to select progeny pre- and post-treatment of IVM treated adults. NGS techniques were employed to identify regions of genetic differentiation between these sampled populations. A new draft reference genome, of similar size to the published genome, PRJNA72569, but one tenth as fragmented was used for analysis of reads. It was expected that resistance-associated alleles would be enriched in the sample populations following treatment of the adults. Initially Farm 2 individuals were sequenced using ddRAD-Seq (chapter 4). Following this, Pool-Seq was used to identify regions of genetic differentiation between pre- and post-IVM populations sampled from Farms 2 and 3 (chapters 5 and 6). Each farm was individually analysed and, based on the hypothesis of a common genetic locus conferring IVM resistance (based on work in *H. contortus*), farm results were compared.

## 2 Materials and Methods

### 2.1 Parasite material

#### 2.1.1 Laboratory isolates

##### ***MTci2 (Moredun T. circumcincta 2)***

MTci2 was obtained by the Moredun Research Institute (MRI) in 2000 from the Central Veterinary Laboratories, Weybridge and is susceptible to most anthelmintic classes. However, there is the possibility of low levels of resistance to BZ (Skuce et al., 2010).

##### ***MTci5 (Moredun T. circumcincta 5)***

MTci5 is a multi-drug resistant isolate, isolated in 2002 from a UK field population following detection of multidrug resistance on the farm the previous year (Sargison et al., 2001) and is resistant to BZ, LEV and IVM (Bartley et al., 2004; Skuce et al., 2010).

#### 2.1.2 Farm 1

A lowland sheep farm in southeast Scotland was visited between March 2016 and March 2017, every three weeks, with the agreement of the shepherd and knowledge of the presence of anthelmintic resistance. Ethical consent was obtained under the guidance of the University of Glasgow. Farm 1 was used to assess current methods available for monitoring anthelmintic resistance. Results are described in Chapter 3.

#### 2.1.3 Farm 2

A Faecal egg count reduction test (FECRT) was performed on a farm in southwest England in August 2016, coordinated by colleagues from the University of Bristol (Prof E. Morgan, Dr H. Vineer and Miss K. Bull), who provided coprocultured L3 pre- and post-IVM treatment. Farm 2 was used for both ddRAD-Seq and Pool-Seq (Chapters 4, 5 and 6).

### 2.1.4 Farm 3

A FECRT was performed on a farm in southeast Scotland in July 2015, coordinated by colleagues at the MRI and coprocultured L3 were provided for use in this study. Farm 3 was used for Pool-Seq (Chapters 5 and 6).

## 2.2 Sample Collection and Processing

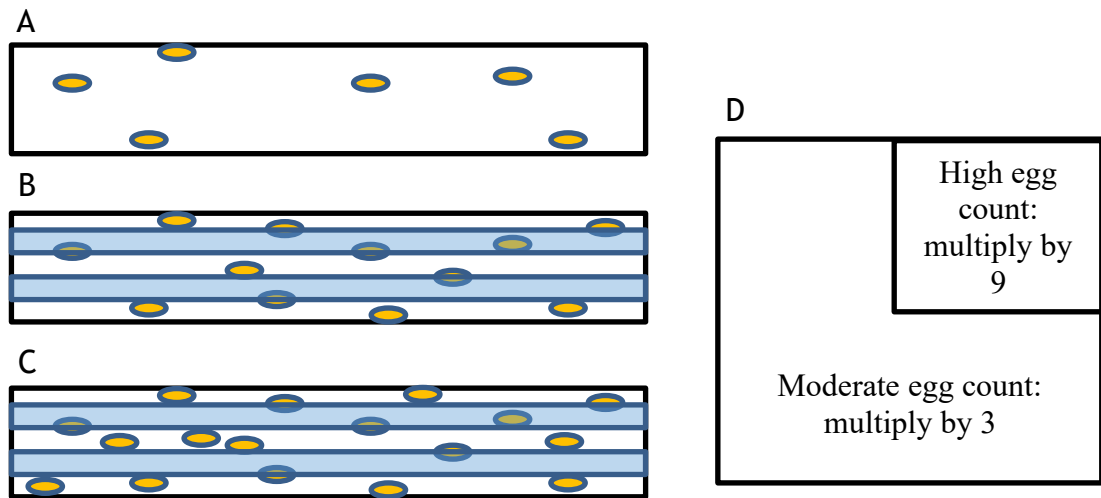
### 2.2.1 Collection of faecal samples

The strongyle population on Farm 1 was monitored over time, in both ewes and lambs. Other than for FECRTs, the identity of individual sheep varied between sampling time points. In line with farm management practices, the same pasture groups were sampled each time. Additional information is provided in 3.3. Starting in March 2016, faecal samples were collected from ten ewes every three weeks until March 2017. From May 2016 until November 2016, samples were also collected from lambs; the first samples were from only four lambs, but after this time point at least ten lambs were sampled every three weeks. Samples were collected as soon as possible after voiding by sheep, and care was taken to minimise contamination with free-living species. Samples were taken quickly to the lab for immediate processing. When egg hatch tests (EHTs) were to be performed, if a delay was likely, such that the estimated start time of the assay was three hours or later from collection, a subsample of faeces was stored anaerobically preventing development of the eggs (Coles et al., 1992). EHTs were always set up the same day.

### 2.2.2 Faecal egg counts

A cuvette method, sensitive to 1 egg per gram (epg) was used to perform strongyle faecal egg counts (FECs) (Christie and Jackson, 1982). Briefly, faeces were weighed and water was added in a ratio of 10 ml to 1 g, following which 10 ml was strained through a coarse sieve (1 mm aperture). After centrifugation at  $234 \times g$  for 2 min, the supernatant was aspirated, and the egg pellet re-suspended in saturated NaCl solution (specific gravity 1.2) with a further centrifugation step as above. The meniscus was isolated using artery forceps and tipped into a cuvette, which was filled with saturated NaCl solution for counting. Strongyle epg and *Nematodirus* epg were quantified (as described in Figure 2.1),

and other species were noted if present. Strongyle eggs were analysed using Minitab (Minitab® 17.1.0, Microsoft). Group arithmetic means were calculated, with 95% confidence intervals (CIs).



**Figure 2.1: Cuvette method for faecal egg counts.** Each of A, B and C represent a cuvette, containing eggs (yellow ovals). Due to the specific gravity of the NaCl solution all eggs floated to the top of the cuvette for counting (Rinaldi et al., 2011). As the equivalent of one gram of faeces was used, each total count is equivalent to the egg. D represents a Miller eyepiece graticule, which can be placed in the microscope eyepiece to enable counting of moderate to high numbers of eggs. A. Low number of eggs present – counted all eggs present within the surface area. For both moderate (e.g. > 50, B) and high (e.g. >300, C) numbers of eggs two rows (shown in blue) were counted using the graticule. Eggs were counted if they were fully within the upper and lower boundaries of the square. In addition, eggs partially covering one of either the upper or the lower boundary were also counted. For moderate egg counts (B), the larger graticule square was used and the sum of both rows multiplied by three. For high egg counts (C), the smaller graticule square was used, and the sum of both rows multiplied by nine.

### 2.2.3 Coprocultures and Baermannisation of L3

Individual faecal samples were pooled separately by age group (ewes or lambs) for culturing. If faeces were too moist, vermiculite was added and faeces formed into small balls for culture at 25 °C for 10 days. Following culture, faeces were immersed in warm water for 6 h before a modified Baermannisation method was used to collect L3. Briefly, the water was poured over a fine muslin mesh, which was placed in a jar containing warm water and left for 24 h, following which the L3 could be collected from the bottom of the jar. L3 were stored in culture flasks in water at 8 °C. Spare L3 were snap frozen in liquid nitrogen and stored at -80 °C.



### 2.2.4 Morphological examination of L3

Larvae were killed with dilute iodine and examined at low magnification, increasing as needed. Sheath tail length was measured, morphological differences in tails and heads were noted (e.g. flat, rounded, tapered, refractile bodies) and intestinal cells counted if needed. If required for identification the entire larval length was measured. This was performed as an initial screen with the help of Mr J. McGoldrick (University of Glasgow) to identify species present for PCR primer selection.

### 2.2.5 Faecal egg count reduction test

A FECRT was performed on Farm 1 in September 2016 (see section 3.3.2.1). Thirty-five, five-and-a-half month old lambs were set aside by the farmer and from these, animals were selected randomly for use in the FECRT, but with the aim of balancing gender. Faeces were collected *per rectum* for FECs pre-treatment, with additional material gathered for further analysis, if voided immediately following treatment. Ten lambs were allocated to the albendazole (BZ) treatment group (5 mg/kg body weight (BW), Albex™ 2.5% w/v SC oral suspension, Chanelle UK), ten lambs to the LEV treatment group (7.5 mg/kg BW, Levacide Drench 3% Oral solution, Norbrook) and twelve lambs to the IVM treatment group (0.2 mg/kg BW, Noromectin® 0.08% w/v Drench Oral Solution, Norbrook). All were weighed individually, using the shepherd's weigh scales (EID weigh crate, Shearwell), and dosed *per os* according to their individual weights using a syringe. Repeat faecal samples were collected on days 7 and 14 post-treatment and faeces transferred anaerobically to the laboratory. The FECR percentage and Bayesian CIs were calculated using R Shiny 'eggCounts' web interface (Wang and Paul, 2017; Wang et al., 2017).

## 2.3 Bioassays

Strongyle eggs were harvested within three hours of faecal collection for immediate use in bioassays as described in 2.2.2, but with the following modifications; emulsified faeces were strained through a series of sieves, (215 µm, 125 µm and 63 µm), with eggs collected from a 38 µm sieve.

Centrifugation steps were performed as described above, and the meniscus was

poured back onto a small, 38 µm sieve. Eggs were washed with tap water to remove salt and collected. A 200 µl aliquot of egg suspension was counted to determine the total number of eggs/ml. The egg concentration was adjusted to 1 egg/µl for the EHT and 2 eggs/µl for the larval development test (LDT). Bioassay results are described in section 3.3.2.2.

### 2.3.1 Egg hatch test

An EHT was set up to measure BZ efficacy, essentially as described previously (Coles et al., 2006). Stock solutions of thiabendazole (TBZ) (Sigma) were prepared by diluting 1 mg/ml of TBZ/DMSO solution into DMSO and stored in opaque tubes at room temperature. Briefly, a 24 well plate was set up with a range of final TBZ concentrations as follows: 0.05 µg/ml, 0.1 µg/ml, 0.2 µg/ml, 0.3 µg/ml, 0.4 µg/ml, 0.5 µg/ml, 1 µg/ml and 0.5% DMSO control (Sigma Aldrich). To each well was added: 10 µl of TBZ/DMSO solution, 1890 µl deionised water (Acros Organics) and 100 µl of egg suspension, giving approximately 100 eggs/well.

EHTs were performed throughout the season on Farm 1, with three replicate wells for each concentration tested. In addition, eggs collected as part of the September BZ and IVM FECRTs on Farm 1 were tested as follows. Four EHTs were performed using pre-treatment samples for each of the BZ and IVM treated lamb groups. For each of the eight tests, all concentrations were repeated in triplicate, when sufficient eggs were available. Plates were incubated in a humid environment at 25 °C for 48 h and the number of unhatched eggs and L1 recorded at 48 h.

On day 14 post-treatment a reduced EHT was performed for the BZ treated population as follows: all TBZ concentrations were incorporated but a reduced quantity of eggs restricted replicate numbers. Two replicates were carried out for each of 0.05 µg/ml, 0.1 µg/ml and 0.2 µg/ml. The rest were single wells. Due to a paucity of eggs post-IVM treatment of lambs, it was not possible to perform a full EHT and only two replicates of the 'definitive dose' wells (0.1 µg/ml) were included.

The  $ED_{50}$ , the effective dose at which 50% of larvae fail to hatch, was calculated using a binomial (probit) general linear model (GLM) in R Studio (version 1.1.383 - © 2009-2017 RStudio, Inc; R version 3.4.3 © 2017). All data were first corrected for the percentage hatch in the DMSO control wells and the TBZ concentrations were  $\log_{10}$  transformed. If the  $ED_{50}$  was greater than 0.1  $\mu\text{g}/\text{ml}$  TBZ, then the sample population was considered resistant to BZ (Coles et al., 1992; von Samson-Himmelstjerna et al., 2009a). The script used for the GLM is available on GitHub (<https://github.com/SheepwormJM>).

### 2.3.2 Larval development test

The LDT was performed at the time of the September FECRT on strongyle populations obtained pre- and post-IVM treatment (day 0 and day 14) and pre-BZ treatment. In total, two tests were performed for each FECRT pre-treatment (total four), and one test was performed post-IVM treatment. Insufficient eggs were recovered post-BZ treatment to allow a LDT to be carried out. A method adapted from Varady et al. (1996) was used to ascertain the ability of larvae to develop in the following concentrations of IVM: 0.57 nM, 1.14 nM, 2.29 nM, 4.54 nM, 9.14 nM, 18.29 nM and 36.58 nM, water only (control), and 2% DMSO (control). Pre-treatment each concentration was included in triplicate in each assay, but post-IVM treatment each concentration was included only twice due to fewer eggs being available. Stock solutions of IVM (22,23 dihydroavermectin B1, primarily  $IVM_{B1a}$ , catalogue number I8898, Sigma Aldrich) were prepared by diluting a 0.001 M IVM/DMSO solution into DMSO and were stored in opaque tubes at room temperature. Briefly, a 48 well plate was set up, each well containing 6  $\mu\text{l}$  of IVM/DMSO solution, 239  $\mu\text{l}$  deionised water and 5  $\mu\text{l}$  of a highly concentrated, reconstituted, lyophilised *E. coli* OP50 as a food source for the developing larvae (Heim et al., 2015). 50  $\mu\text{l}$  of strongyle eggs were added (concentration of 2 eggs/ $\mu\text{l}$ ), giving approximately 100 eggs/well. No antimicrobials were used. Wells surrounding those used for the assay were filled with deionised water and the plate sealed with an adhesive plastic seal (Sigma Aldrich). The assay was incubated for 7 days at 25 °C in a humid environment. Eggs, L1/L2 and L3 larvae were counted. Those with obvious failure (e.g. no development, fungal or bacterial overgrowth) were discarded. Following correction for development in the control wells, an  $ED_{50}$  for each sample was

calculated using a GLM constructed in R Studio as above. The script used for the GLM is available on GitHub (<https://github.com/SheepwormJM>).

## 2.4 Molecular Biology

### 2.4.1 Lysate protocol

Lysates were made from individual strongyle eggs, L1 or L3 in 96 well plates. Briefly, a master mix per 100 wells was prepared as follows: 1000  $\mu$ l DirectPCR Lysis Reagent (Cell; Viagen Biotech), 50  $\mu$ l 1 M DTT (Thermofisher Scientific) and 10  $\mu$ l Proteinase K (Fungal; Invitrogen), which had been reconstituted at 100 mg/ml in DEPC treated water (Ambion). 10  $\mu$ l was dispensed per well and a single egg or larva added to each in  $\leq$  1  $\mu$ l of water/PCR Lysis Reagent. After incubation at 60 °C for 2 h, followed by 85 °C for 45 min to denature the proteinase K, lysates were aliquoted in 1:20 dilutions using DEPC treated water. All lysates were stored at -80 °C.

### 2.4.2 Speciation of strongyles by PCR

Individual eggs, L1 or L3, lysed as described in 2.4.1, were identified by PCR to species level using the ITS2 region (Wimmer et al., 2004; Redman et al., 2008; Burgess et al., 2012; Bisset et al., 2014). Two PCR methods were employed: single species PCR and a multiplex PCR. Briefly, single species PCR was performed using GoTaq Flexi polymerase (Promega) according to the manufacturer's instructions. The total volume of each reaction was 12.5  $\mu$ l of which 1  $\mu$ l was 1:20 DNA lysate. Primers designed to amplify the ITS2 region and specific to the species of interest were used. Table 2.1 shows primer sequences with their original reference, expected amplicon size, and  $T_A$  °C. All PCRs were carried out using the following protocol: denaturation at 94 °C for 2 min, then 35 cycles each of 94 °C 30 s,  $T_A$  °C 30 s and 72 °C 30 s, with a final extension step of 72 °C 10 min.

**Table 2.1: Primer sequences used to speciate strongyle larvae.**

Species	Primer name	Sequence	T <sub>A</sub> (°C)	Product size (bp)	Reference
<i>Teladorsagia circumcincta</i>	Tc_ITS2	F: ATACCGCATGGTGTGTACGG	52	421	(Burgess et al., 2012)
		R: CAGGAACGTTACGACGGTAAT			(Burgess et al., 2012)
<i>Trichostrongylus vitrinus</i>	Tv_ITS2	F: AGGAACATTAATGTCGTTACA	52	100	(Wimmer et al., 2004)
		R: CTGTTTGTCTGAATGGTTATTA			(Wimmer et al., 2004)
<i>Haemonchus contortus</i>	Hc_ITS2	F: GTTACAATTCATAACATCACGT	50	321	(Redman et al., 2008)
		R: TTTACAGTTTGCAGAACTTA			(Redman et al., 2008)
Generic ITS2	ITS2GF	F: CACGAATTGCAGACGCTTAG	54	370-398	(Bisset et al., 2014)
Generic ITS2	ITS2GR	R: GCTAAATGATATGCTTAAGTTCAGC	54		(Bisset et al., 2014)
<i>Chabertia ovina</i>	ChovFd2	F: CAGCGACTAAGAATGCTTTGG	54	115/117	(Bisset et al., 2014)
<i>Cooperia curticei</i>	CocuFd3	F: TAATGGCATTGTCTACATTGGTTC	53	252	(Bisset et al., 2014)
<i>Oesophagostomum venulosum</i>	OeveRv1	R: CGACTACGGTTGTCTCATTTCA	54	323-329	(Bisset et al., 2014)
<i>Trichostrongylus axei</i>	TraxFd2	F: GATGTTAATGTTGAACGACATTAATATC	52	186	(Bisset et al., 2014)

Multiplex PCR was also performed, adapted from a protocol optimised by Bisset et al. (2014). Briefly, each 12.5 µl reaction contained: 2.5 µl 5X GoTaq Green buffer, 1.25 µl MgCl<sub>2</sub> 25 mM, 0.06 µl GoTaq Flexi polymerase 5 U/µl (Promega), 0.25 µl dNTPs 10 mM each (NEB Inc.), 0.15 µl Generic Forward primer 10 µM, 0.15 µl Generic Reverse primer 10 µM, 0.25 µl *Trichostrongylus axei* Forward primer 10 µM, 0.35 µl *Cooperia curticei* Forward 3 primer 10 µM, 0.1 µl *Oesophagostomum venulosum* Reverse 1 primer 10 µM, 0.15 µl *Chabertia ovina* Forward 2 primer 10 µM (Eurofins Genomics, Table 2.1), and 6.29 µl DEPC treated water (Ambion). To this 1 µl of 1:20 DNA lysate was added. Touchdown PCR conditions were: denaturation at 94 °C for 2 min, then 12 cycles each of 94 °C 15 s, 60 °C, decreasing by 0.5 °C each cycle, to 54.5 °C for 15 s and lastly 72 °C 30 s. Next, 25 cycles each of 94 °C 15 s, 54 °C 15 s, 72 °C 30 s, with a final extension step of 72 °C for 7 min. Products were visualised with SafeView (NBS Biologicals) on 2% agarose gels for single species PCR and 2.5% agarose gels for multiplexed PCR.

Any unidentified strongyles (i.e. positive on the multiplex PCR as a strongylid nematode) were amplified with Phusion High-Fidelity Polymerase (NEB) using the generic ITS2 primer set (Bisset et al., 2014) and sequenced by Eurofins Genomics. A BLAST search using NCBI was performed for identification (Altschul et al., 1990). Speciation results were analysed in Microsoft Excel and species proportions, determined by PCR, were used to estimate individual *T. circumcincta* egg counts pre- and post-treatment. A 2-sample proportion test was performed in Minitab to compare speciation results pre- and post-treatment of the FECRT performed on Farm 1.

### 2.4.3 Microsatellite PCR and Analysis

Ten microsatellites were amplified by PCR. Briefly, each 20 µl reaction contained: 4 µl 10% Tween 20, 1.8 µl 11.1X buffer, 1 µl Forward primer, 1 µl Reverse primer (Eurofins Genomics), 0.1 µl of GoTaq Flexi polymerase (Promega), 11.1 µl DEPC treated water (Ambion) and 1 µl 1:20 DNA lysate. Table 2.2 shows primer sequences with their original reference, expected amplicon size range, microsatellite repeat sequence and T<sub>A</sub> °C. The 11.1X buffer was made in house, and contained 334 µl 2M Tris HCl (pH 8.8), 166 µl 1M Ammonium sulphate, 67 µl 1M MgCl<sub>2</sub>, 7.2 µl 100% 2-mercaptoethanol, 6.8 µl 10 mM EDTA (pH

8), 150  $\mu$ l 100mM dATP, 150  $\mu$ l 100 mM dCTP, 150  $\mu$ l 100mM dGTP, 150  $\mu$ l 100 mM dTTP and 170  $\mu$ l 10 mg/ml bovine serum albumin. Aliquots were stored at -20 °C. Thirty *T. circumcincta* L3 were tested per field sample population. PCR conditions were as follows: 94 °C for 2 min, then 40 cycles each of 94 °C for 15 s,  $T_A$  °C for 45 s and 72 °C for 1 min and a final extension step of 72 °C for 15 min. Products were visualised with SafeView (NBS Biologicals) and the brightness of the visualised band compared with the same volume of a 50  $\mu$ g/ml Quick-Load® 100 bp DNA ladder (NEB). This allowed a qualitative estimation of DNA quantity for subsequent dilution steps. PCR products were wrapped in tinfoil and stored at 4 °C. Products were diluted and multiplexed as required (Table 2.2, footnote) and sent to the MRC PPU DNA Sequencing and Services laboratory, University of Dundee. An Abi Prism 3730xl Genetic Analyser (Applied Biosystems), with 50 cm capillaries was used to quantitatively measure the size of each amplified fragment. GeneScan™ 500 ROX™ dye size standard (Applied Biosystems) was selected.

Results were analysed using GeneMapper™ Software v4.0 (Applied Biosystems) to automatically call peaks, using information including the expected tandem repeat size and fragment size range. Following this, peaks were manually checked and data exported to Microsoft Excel. TANDEM v1.09 was used to assign raw allele calls to tandem repeat size bins (Matschiner and Salzburger, 2009). Following this, data were converted into Arlequin format using the software Convert v1.31 (Glaubitz, 2004). Arlequin v3.5.2.2 was used to statistically analyse data for genetic relatedness between samples (Excoffier et al., 2005). Settings used can be found in Appendix 1. Results are described in section 3.3.3.

**Table 2.2: Primer sequences for microsatellites.**

Primer name	Repeat sequence	Sequence	Label	T <sub>A</sub> (°C)	Product size (bp)	Reference
MTG15 <sup>a</sup>	[GT] <sub>6</sub> GC[GT] <sub>6</sub> GGGTTTGT	F: TGCAAGGAAACTGCTAAGAAGGAG	5' FAM	58	233-275	(Grillo et al., 2006)
		R: ATCATGGAACCTTGATACCGCAAG				
MTG67 <sup>a</sup>	[CT] <sub>14</sub>	F: CAAGTCGTTTAGGCACGTCTGG	5' HEX	58	172-192	(Grillo et al., 2006)
		R: CAGGGCGGAACCCAATTGATCG				
MTG68 <sup>b</sup>	[ACA] <sub>2</sub> [ACC] <sub>2</sub> TCG[ACA] <sub>4</sub> ACTA CAACCACAACACTACG[ACAACC] <sub>2</sub>	F: ATCACCAGGCGGCTGCTACG	5' FAM	50	420-453	(Grillo et al., 2006)
		R: CGAAAAGTAGAGTATGAGC				
MTG73 <sup>b</sup>	[TGC] <sub>2</sub> TGTTGC[TGT] <sub>2</sub> [TGA] <sub>3</sub> T G[TGA] <sub>2</sub> TAA[TGA] <sub>2</sub> AAATA[TGA] <sub>2</sub>	F: CCTTGTATAAATTCGAAG C	5' HEX	45	148-157	(Grillo et al., 2006)
		R: GTAGTAGTGATTAACCTCCG				
Tc13604 <sup>a</sup>	[CAGGTA] <sub>12</sub> TAGGTACAGGCACAG	F: CGATAAATGGTATTATCTG	5' HEX	50	250-386	(Redman et al., 2015)
		R: GCTGCTATTAGAGGATAT				
Tc2066 <sup>c</sup>	[GGCGAGTA] <sub>9</sub> GGCGATTAGGCGA GAAGGCGGTAGGTGAGTAGGCGAG	F: GAGCAACGACTGAACCTCAC	5' HEX	50	189-309	(Redman et al., 2015)
		R: GCTGGAAGCATATTCTGCGC				
Tc22274 <sup>d</sup>	[TGTA] <sub>17</sub> TGTCTGAATGT	F: ACAAAGTGCTCAAGTTAG	5' HEX	54	189-329	(Redman et al., 2015)
		R: GGGGGTTCTATATACAGTA				
Tc2467 <sup>d</sup>	[TTTA] <sub>15</sub> TCTATT	F: AACGCTTTGAACCGTGTCGG	5' FAM	54	149-224	(Redman et al., 2015)
		R: GCTGCCACATCAGCTTAGA				
Tc4504 <sup>b</sup>	[ACAT] <sub>13</sub> ACAGATACAGACA	F: TTATCACACCACTTCATT	5' FAM	50	228-276	(Redman et al., 2015)
		R: GTCTTTAAACGCTAAATA				
Tc7989 <sup>c</sup>	[GTCT] <sub>17</sub> GTCCGTTTGTGTG	F: GATCTCACGTAATGAA	5' FAM	50	133-232	(Redman et al., 2015)
		R: CTATTGAATGTCGTACAG				

Primers were multiplexed following PCR for fragment analysis as follows (dilutions shown in brackets); <sup>a</sup>MTG15 (1:20), MTG67 (1:20) and Tc13604 (1:20), <sup>b</sup>MTG68 (1:5), MTG73 (7:10) and Tc4504 (1:10), <sup>c</sup>Tc2066 (1:2) and Tc7989 (1:2), <sup>d</sup>Tc22274 (1:50) and Tc2467 (1:50).



#### 2.4.4 Pyrosequencing

Using a method adapted from Skuce et al. (2010), BZ resistance associated SNPs were pyrosequenced in *T. circumcincta* individuals from Farm 1 (30 individuals per sample - see section 3.3.2.3). Eggs and L1, obtained from EHTs were used on day 14 post-BZ and post-IVM FECRTs. At all other time points L3 were used. Two regions of the  $\beta$ -tubulin isotype-1 gene (one region including codons 200 and 198, and the other including codon 167) were amplified using the Pyromark PCR kit (Qiagen) according to the manufacturer's instructions. Briefly, a 25  $\mu$ l reaction was set up using 12.5  $\mu$ l 2X Pyromark Master Mix, 0.5  $\mu$ l of each primer, 10.5  $\mu$ l DEPC treated water (Ambion) and 1  $\mu$ l of diluted DNA lysate. PCR conditions were: denaturation at 95 °C for 15 min, then 45 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, with a final extension step of 72 °C for 10 min. Two positive controls were included each time, one MTci2 (sensitive) and one MTci5 (resistant). Pyrosequencing was performed on a Biotage Q96 ID instrument (Qiagen) according to the manufacturer's protocol with pyrograms evaluated manually. The sequences analysed were: 'CAGWAYGTYDMRTCGG' for codons F200Y and E198A/L, and 'TCATWCTC' for codon F167Y. Primers used are shown in Table 2.3. A 2-sample proportion test was performed in Minitab to compare pyrosequencing results from samples collected pre- and post-anthelmintic treatments on Farm 1.

**Table 2.3: Primer sequences for pyrosequencing of the BZ resistance associated SNPs in  $\beta$ -tubulin isotype-1.**

Primer name	Sequence	T <sub>A</sub> (°C)	Label	Product size (bp)	Reference
Btub_SK_200	F: ACCTTACAATGCCACTCTTTCTG	55	Biotin	97	(Skuce et al., 2010)
	R: GCGGAAGCAGATATCGTACAG				(Skuce et al., 2010)
	SEQ: RGAGCYTCATTATCGATR				(Skuce et al., 2010)
Btub_SK_167	F: GCATTCTTTGGGAGGAGGTA	55		122	(Skuce et al., 2010)
	R: TGCACCTCGAGAACCTGTACATA		Biotin		(Skuce et al., 2010)
	SEQ: CGGATAGAATCATGGCT				(Skuce et al., 2010)

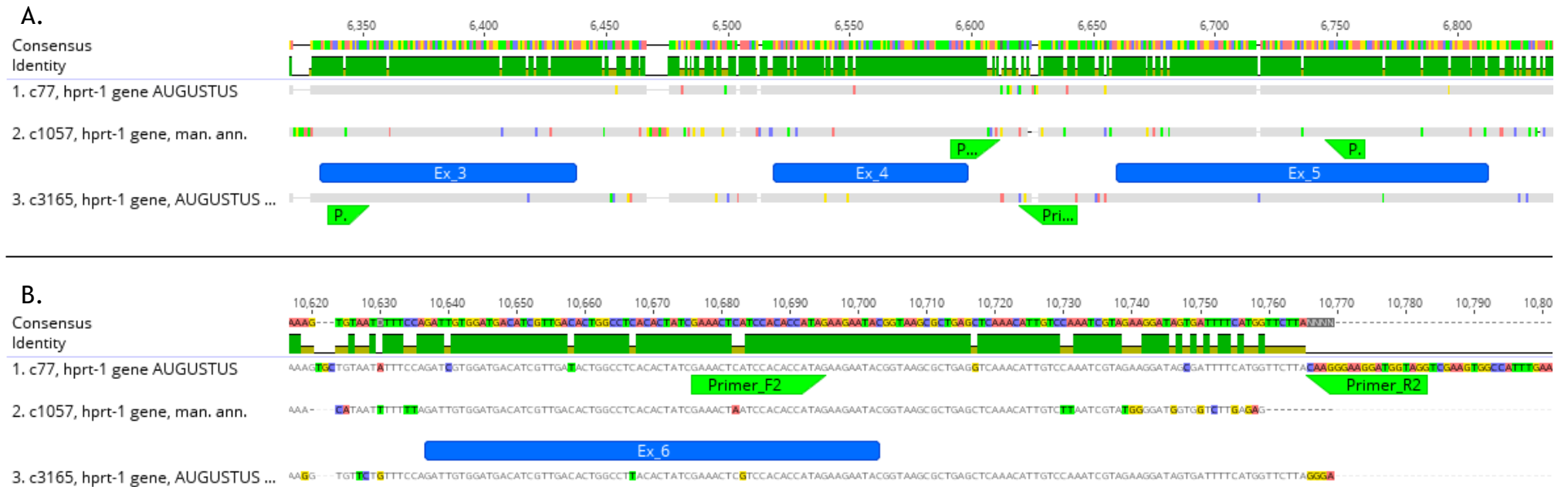
For each amplified region, a forward and reverse primer pair was used to amplify a fragment containing the relevant SNPs. The sequencing primer (SEQ) was used for the pyrosequencing reaction.

### 2.4.5 *hprt-1* allele specific PCR

Three primer pairs were manually designed using sequences aligned in Geneious v9.1.8, each pair specific to one of three potential alleles of the predicted homologue of *hprt-1* (Figure 2.2). Using GoTaq Flexi polymerase (Promega) according to the manufacturer's instructions, each primer pair was tested on 96 individual MTci2 L3. The total volume of each reaction was 12.5  $\mu$ l of which 1  $\mu$ l was 1:20 DNA lysate. Table 2.4 shows primer sequences with their expected amplicon size, and  $T_A$  °C. All PCRs were carried out using the following protocol: denaturation at 94 °C for 2 min, then 35 cycles each of 94 °C 30 s,  $T_A$  °C 30 s and 72 °C 30 s, with a final extension step of 72 °C 10 min. Products were visualised with SafeView (NBS Biologicals) on 2% agarose gels. Rationale and results are provided in section 6.3.7.

**Table 2.4: Primer sequences for *hprt-1* alleles.**

Primer name	Sequence	$T_A$ (°C)	Product size (bp)
Tci2_hprt-1_c77	F: GAAACTCATCCACACCATAG	50	108
	R: CCTACCATCCTTCCCTTG		
Tci2_hprt-1_c1057	F: AACACGTTGGTTAGTTCAATG	50	166
	R: ACCTTTCCAGAAGTGGC		
Tci2_hprt-1_c3165	F: TACCGGTAAGCGCTTTTCG	50	296
	R: TTAAACTAGACGAAAACGACAG		



**Figure 2.2: Primer positions to amplify *hprt-1* gene copies.** The genes were extracted from the Tci2 contigs and aligned in Geneious v9.1.8. A. and B. Three primer sets were manually designed to amplify each gene copy individually. Primer properties were checked using OligoCalc (Kibbe, 2007) and the gene sequences were searched in Geneious v9.1.8 for multiple primer alignment. Exons are annotated in blue. Primer positions are indicated by lime green annotations. Each sequence is shown on a separate line, with a consensus sequence displayed at the top. Sequence similarity is indicated by a coloured bar plot between the consensus and sample sequences. Where there is complete agreement between sequences this is shown in green, with reduced height and varying colour from yellow to red for decreased consensus. Differences between sample sequences are highlighted by colours for different basepairs, while regions with no difference to another sequence (or no overlapping sequence) are shown in grey.

## 2.5 Genomic DNA Library Preparation for Next Generation Sequencing: ddRAD-Seq and Pool-Seq

Next generation sequencing techniques were used as part of a genome wide association study investigating IVM resistance in UK field populations of *T. circumcincta*. Information about these techniques and the results obtained are presented in Chapters 4 (ddRAD-Seq), 5 and 6 (Pool-Seq).

### 2.5.1 Whole Genome Amplification of DNA for ddRAD-Seq

Thirty pre- and 32 post-IVM FECRT *T. circumcincta* L3 from Farm 2 were selected for whole genome amplification (WGA) and ddRAD-Seq. A maximum of eight larvae were processed at any one time. To decrease variation due to batch error, pre- and post-IVM individuals were selected in roughly equal numbers for each batch of samples prepared. For WGA the polymerase Phi-29 was used, capable of producing long stranded DNA molecules (>10 kb) via multiple displacement amplification at a constant temperature. Briefly, 5 µl worm lysate prepared as described in 2.4.1 was combined with 5 µl 1.3 M Trehalose and heated at 95 °C for 3 min. On ice, a master mix was prepared containing 29 µl REPLI-g buffer, 1 µl REPLI-g polymerase (REPLI-g Single Cell Kit, Qiagen) and 10 µl 1.3 M Trehalose per sample and 40 µl of this was transferred to each sample. The reaction was heated at 30 °C for 16 h before denaturing Phi-29 at 65 °C for 10 min.

The WGA DNA was purified using Agencourt AMPure XP beads (Beckman Coulter Life Sciences) as per the manufacturer's protocol, with a ratio of 1.8X beads to DNA volume. DNA was quantified using the Qubit™ 2.0 Fluorometric quantitation protocol (Invitrogen). WGA was assessed qualitatively using primers designed to amplify six core eukaryotic genes. Each gene was predicted to be located on a different chromosome of each of the six nematode chromosomes based on synteny analysis between *C. elegans* and *H. contortus*. One-to-one *T. circumcincta* homologues of the *H. contortus* genes were identified using the WormBase ParaSite BLAST (Howe et al., 2016; Howe et al., 2017). Primers were designed using Geneious® v9.1.8 (<https://www.geneious.com>) and are shown in Table 2.5. Primers were tested using individual larval lysates and WGA DNA. A qualitative difference in the relative brightness of amplicon bands was obtained

and WGA DNA of each individual was assessed in this way. Briefly, each PCR was performed using GoTaq Flexi polymerase (Promega) according to the manufacturer's instructions. The total volume of each reaction was 12.5  $\mu$ l of which 1  $\mu$ l was WGA DNA, while a further sample contained 1  $\mu$ l of 2:22 DNA lysate as a negative control (of WGA failure). A known 1:10 WGA DNA sample was used as a positive control of WGA success. As always, a negative control with no DNA was included. The PCR conditions were: 94 °C 2 min, then 35 cycles of 94 °C 30 s, 55 °C 30 s and 72 °C 25 s, followed by an extension step at 72 °C for 10 min. PCR products were visualised on a 2% agarose gel with SafeView (NBS Biologicals).

**Table 2.5: Primer sequences for core eukaryotic genes.**

Primer name	Sequence	T <sub>A</sub> (°C)	Product size (bp)	<i>C. elegans</i> chromosome
<i>copb-1</i>	F: CTCCATGCGGCTACGATCA	55	238	II
	R: TGCAACGAATGGAGAATTGTC			
<i>dyf-7</i>	F: CCAAACCTACCGAACTGCAGA	55	246/247	X
	R: CCACCAGTCGATGCCAAC			
<i>gspd-1</i>	F: GAGCTTGATCTCACTTACAATG	55	345	IV
	R: GCGTCAGGTAAACGAGTATC			
<i>hmgs-1</i>	F: GCGTGAAGACAGAGTTGATG	55	318	V
	R: GCTTTTCCATTGAGTAGTTTGC			
<i>mel-32</i>	F: CTCGAAATTTGGACTATGCTC	55	272	III
	R: GATCAGTGCTCCACGAGG			
<i>smo-1</i>	F: AGATAATGGCAGCGAAGCAC	55	265	I
	R: AATGAAGAGACGGCAACTCC			

### 2.5.2 Double digest of WGA DNA

In order to reduce GC bias (Aird et al., 2011; Dabney and Meyer, 2012), the restriction endonucleases MluCI and NlaIII were chosen. 900 ng of WGA DNA was digested using MluCI and NlaIII restriction endonucleases (NEB). Briefly, each reaction contained 9  $\mu$ l 10X Cut Smart buffer (NEB), 9  $\mu$ l MluCI, 9  $\mu$ l NlaIII and 900 ng WGA DNA with the total volume made to 90  $\mu$ l with DEPC treated water (Ambion). The digestion was completed at 37 °C for 4 h. The double digest was purified and quantified as for 2.5.1.

### 2.5.3 Adaptor ligation

Barcoded adaptors, designed by Peterson et al. (2012), and kindly supplied by the Wellcome Sanger Institute (WSI), were ligated to 250 ng of digested WGA

DNA to ensure identification of individuals when pooled. Only if both a P1 and a P2 adaptor, each specific to either MluCI (AATT) or NlaIII (CATG) cut-sites, were ligated to a DNA fragment would sequencing occur (Peterson et al., 2012). Each adaptor ligation reaction contained 4 µl 10X T4 DNA ligase buffer, 0.7 µl T4 DNA ligase 2,000,000 U/ml (NEB), 5 µl P1 adaptor, 5 µl P2 adaptor (Peterson et al., 2012), 250 ng digested WGA DNA and the total made to 40 µl with DEPC treated water (Ambion). Samples were heated at 23 °C for 30 min then at 65 °C for 10 min. Reactions were cooled by 2 °C every 30 s until the reaction reached 19 °C. DNA was stored at -20 °C prior to pooling of individuals. Each pool included 15 pre- and 16 post-IVM L3. Agencourt AMPure XP beads (Beckman Coulter Life Sciences) were used as in 2.5.1 to purify and gradually reduce the overall volume of each library to 40 µl.

#### **2.5.4 Size selection, Illumina Library Prep and DNA Sequencing**

Genomic DNA libraries were sent to the WSI, where 225-325 bp fragments were size selected using a Pippin Prep (Sage science). Illumina adapters were ligated and eight rounds of KAPAHiFi (KAPA Biosystems) PCR were performed, with post-PCR SPRI bead clean-up and quantification by qPCR. Genomic libraries were sequenced with 125 bp paired-end reads on an Illumina HiSeq 2000v4.

#### **2.5.5 Preparation of Genomic DNA Libraries for Pool-Seq**

L3 pre- and post-IVM treatment from Farms 2 and 3 were also prepared for Pool-Seq. For each of the sample populations pre- and post-IVM treatment, neat lysate material prepared as in 2.4.1, from 91 individually lysed L3 was pooled for extraction of gDNA. Farm 2 had already been used for ddRAD-Seq, so a maximum of 5 µl from each L3 lysate was available (half total L3 DNA). Farm 3 had 9 µl per L3 lysate available (~90% L3 DNA). To keep the protocol consistent between farms, Farm 3 gDNA was extracted in two parts; first 5 µl was pooled from each L3, then the remainder of the lysate was pooled separately, providing a technical replicate for each of the pre- and post-IVM sample populations from Farm 3.

Genomic DNA was extracted as follows. Briefly, 15 µl of 10 mg/ml RNase A (Thermofisher) was added to the pooled lysates, before incubating at 37 °C for

10 min using a Techne DRI-BLOCK® DB-2A. Next, 792 µl of UltraPure™ phenol:chloroform:IAA (25:24:1) (Invitrogen) was added and the mixture shaken vigorously for 15 s, before being incubated at room temperature for 5 min, turning every 1 min. The solution was centrifuged at 14,000 x g for 15 min at room temperature and the aqueous supernatant transferred to a new Eppendorf tube. To remove any remaining phenol, 300 µl of chloroform:IAA (Sigma Aldrich) was added and the suspension inverted, before a second centrifugation step at 14,000 x g for 15 min. The aqueous supernatant was transferred to a new Eppendorf tube, and measured. To this was added 0.1X vol 3M NaOAc pH 5.5, 3X vol 100% Ethanol Absolute Electran Molecular Biology grade (VWR Chemicals, at room temperature) and 2 µl of 5 mg/ml glycogen (Invitrogen), before storage at -80 °C to precipitate the gDNA. Following precipitation, the mixture was centrifuged at 14,000 x g at 4 °C for 45 min. The supernatant was aspirated and the pellet resuspended in 500 µl 70% ethanol. The gDNA was centrifuged at 14,000 x g at 4 °C for 5 min and the supernatant carefully removed. A final brief centrifugation step of 14,000 x g at 4 °C for 30 s was applied, any remaining ethanol aspirated and the pellet air dried before resuspension in 10 µl of EB buffer (Invitrogen). The gDNA was stored at 4 °C before being sent to the WSI. Libraries were prepared and 75 bp paired-end sequencing was performed on an Illumina HiSeq 4000.

## 2.6 Bioinformatics: Reference genomes

At the end of October 2017, a draft *T. circumcincta* genome of length 685 Mb, N50 288.5 kb, in 8025 contigs and with the largest contig of size 2.08 Mb had been assembled by Dr. S. Doyle at the WSI (Tc\_171026.fa). The genome was assembled using MTci2 worms and, throughout the rest of the thesis this genome is referred to as the 'Tci2 genome'. Generated using PacBio methodology (Eid et al., 2009), corrected by Illumina reads, the genome was a significant improvement on the highly fragmented PRJNA72569 genome. This second genome is based on a susceptible NZ isolate, published on WormBase ParaSite, and is in 81,734 contigs (Choi et al., 2017). However, the PRJNA72569 genome (701 Mb, N50 46.8 kb), had the advantage of protein level gene annotations, although some were incomplete or incorrect. The Tci2 genome has been used as the main reference genome in this study.



## 2.7 Bioinformatics: ddRAD-Seq

Farm 2 was used for ddRAD-Seq. Following validation of the method, a pairwise comparison between the pre- and post-treatment samples was performed to identify regions of genetic differentiation. Bioinformatics methods are detailed below. Note that scripts used are included in Appendix 2 when relevant, and are also available on GitHub (<https://github.com/SheepwormJM>). Raw data can be accessed via ENA using accession numbers ERS1770619 and ERS1770620.

### 2.7.1 Cleaning and demultiplexing of ddRAD-Seq reads

Fastqc (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v0.9 (Ewels et al., 2016) were used to visually assess the quality of the sequencing data. Reads were de-multiplexed and sequencing barcodes removed by Stacks v2.0beta2 *process\_radtags* (Catchen et al., 2013; Rochette et al., 2019), resulting either in 120 bp or 125 bp read lengths. Each base of each read had been previously assigned an Illumina Phred score. This is the measure of the probability (P) that the base called is incorrect and is calculated as:

$$Q = -10[\log_{10}(P)] \text{ (Illumina, 2011).}$$

Any reads which decreased in quality below a Phred score of 10, using a sliding window of 15% of the read length, were discarded by *process\_radtags*. From this point forwards, only paired-end reads with an average base quality >10 were used. The quality of the reads per individual was assessed using Fastqc and MultiQC.

### 2.7.2 Alignment of ddRAD-Seq reads to Tci2 genome

Reads were aligned to the Tci2 genome using Burrows Wheeler Aligner (BWA-MEM) v0.7.17, (Li and Durbin, 2009; Li, 2013), after first splitting the reads into batches using a custom script written by Dr. S. Doyle, in order to increase processing speed. Command line parameters included -Y, to use soft-clipping for supplementary alignments and -M to call split reads as 'secondary'. Duplicates were not marked, as these are difficult to identify for ddRAD-Seq. SAM files generated were merged and sorted, before conversion to BAM output files using SAMtools v1.3 (Li et al., 2009). Read alignment was analysed using SAMtools v1.3

*flagstats* and SAMtools v1.3 *stats*. MultiQC v0.9 was used to visualise alignment success. Paired, two tailed Student's t-tests were performed using Microsoft Excel, to check for statistical differences in read alignment between pre- and post-treatment samples, as described in section 4.3.3.

### 2.7.3 Formation and filtering of RADloci

Aligned reads were provided as input to the Stacks v2.0b module *gstacks*, which groups reads by similarity into 'stacks' of haplotypes and then RADloci (Catchen et al., 2013; Rochette et al., 2019). RADloci are genomic regions between two restriction endonuclease cut-sites. Haplotypes within and between individuals, predicted to be sequenced from the same genomic region, are labelled with identical RADlocus ID numbers. The *gstacks* SNP calling model, based on the low coverage model of Maruki and Lynch (2017), was requested to only retain variant sites and call genotypes when calculated to have a probability of being incorrect of  $\leq 0.001$ . Following genotyping and variant calling, a *populations* analysis was performed. This module was initially run using all individuals, however following kinship analysis (2.7.4), two individuals (one pre- and one post-treatment) were removed and the *populations* module repeated for further analyses. The rationale for this is explained in section 4.3.5. RADloci retained were present in both the pre- and post-treatment sample populations and were sequenced in at least 80% of individuals in each sample ( $-p2$ ,  $-r0.8$ ) (Paris et al., 2017). Variant sites within retained RADloci were further analysed if the minor allele frequency for the genotype was at least 0.048 overall, and the maximum heterozygosity observed did not exceed 70% ( $--min\_maf$  0.048,  $--max\_obvs\_het$  0.7) (Rochette and Catchen, 2017). Results are described in section 4.3.4.

### 2.7.4 Population structure and Kinship analysis

The software Adegenet v2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011) was used to assess the genetic structure of the sample populations. From the Stacks plink output files, a plink.raw file was generated using PLINK v1.90b3v (Chang et al., 2015) and loaded as a genlight file in Adegenet v2.1.1 in R Studio (RStudio-Team, 2015). A plot of individuals genotypes (0 - homREF, 1 - het, 2 - homALT) for each variant site was generated. Principal Component Analysis (PCA) was performed using Adegenet and plotted using *ggplot2* v3.2.1 (Wickham, 2016).

Using *vcfR* v1.7.0 (Knaus and Grünwald, 2017), a genlight R file was produced, and sample group information and ploidy added. Using an Unweighted Pair Group Method with Arithmetic Mean (UPGMA, average neighbour method), the package *poppr* v2.8.2 was used to calculate the bitwise distance by bootstrapping (*aboot*, sampling 100 times) (Kamvar et al., 2014; Kamvar et al., 2015). A distance genetic tree, coloured by sample group, was plotted with *ggtree* (Yu et al., 2017).

Variant data from each individual was used to investigate genetic relatedness. The SNPRelate R package v1.12.0 (Zheng et al., 2012) module *Kinship-based Inference for GWAS (KING)* (Manichaikul et al., 2010) was used to determine the proportion of variants that were identical-by-state and hence the familial relatedness between individuals. The *KING-robust* algorithm, not impacted by linkage disequilibrium and able to cope with population stratification, was used to compute kinship coefficients for each pair of individuals, allowing population substructure to be revealed. Results are described in section 4.3.5.

### 2.7.5 Hardy Weinberg analysis

Stacks v2.0b *populations* was used to calculate departure from Hardy-Weinberg Equilibrium (HWE) using an exact test for SNPs (Louis and Dempster, 1987; Engels, 2009) and a Markov Chain for RADloci haplotypes (Guo and Thompson, 1992; Rochette et al., 2019). In addition, VCFtools v0.1.14 was used to assess SNPs for heterozygosity excess or deficit (*--hardy*) (Danecek et al., 2011). Results are described in section 4.3.6.

### 2.7.6 Linkage disequilibrium analysis

Prior to calculating the linkage disequilibrium (LD), the percentage of missing variant sites per individual was calculated using VCFtools v0.1.14 (*--missing-indv*). Six individuals were removed (three from each of the pre- and post-treatment samples) and Stacks v2.0b *populations* repeated with 54 individuals, but requesting the proportion of individuals per site in each population to be equivalent to the original *-r80* value (Catchen et al., 2013; Paris et al., 2017). The new variant output file was provided to VCFtools v0.1.14 to calculate the LD  $r^2$  value between each variant site within contigs (*--geno-r2*) and between

contigs (*--interchrom-geno-r2*). Note that this compares genotypes (homozygous for the reference, heterozygous or homozygous for the alternate allele) rather than phased alleles as haplotype data are not available. The LD decay was calculated and plotted using an R script written by Fabio Marroni (<https://fabiomarroni.wordpress.com/2011/08/09/estimate-decay-of-linkage-disequilibrium-with-distance/>, last accessed Jan 2019). The distance in basepairs to the mean intercontig LD  $r^2$  value (Cutter et al., 2006), was also obtained using this script. Results are described in section 4.3.8.

### 2.7.7 SNP density across the genome

The SNP density per kb was calculated using VCFtools v0.1.14 *--snpden*. Results were calculated in distinct 1 kb windows across the genome. The first 50 contigs were plotted using R to visually assess SNP density. Results are described in section 4.3.9.

### 2.7.8 ddRAD-Seq genetic differentiation analysis

$F_{ST}$  values were calculated for each RADmarker (SNP) using Stacks v2.0b *populations*. Note that  $F_{ST}$  is used throughout this PhD as a description of the degree of genetic differentiation between sample populations. It is particularly used to compare the pre- and post-IVM populations in the ddRAD-Seq chapter (chapter 4) and later for the Pool-Seq chapters (chapters 5 and 6).  $F_{ST}$  is not a true measure of genetic differentiation but is rather a fixation index, describing the degree of fixation of alleles between populations. It ranges from 1 (complete fixation) to 0 (each population has equal proportions of each allele - panmixis).  $F_{ST}$  will be low in the presence of within-population diversity, and therefore the relative difference between  $F_{ST}$  values are usually considered rather than absolute values. In addition,  $F'_{ST}$  and  $D_{EST}$  statistics were calculated for RADloci (Bird et al., 2011), as these can better account for within-population diversity.  $D_{EST}$  is a true measure of genetic differentiation, while  $F'_{ST}$  is a fixation measure which calculates the relative degree of allele fixation between populations given the within-population diversity. For each statistical dataset, RADloci and RADmarkers which had calculated values of genetic differentiation greater than three standard deviations above the mean value were considered of interest and were retained for further study. At the time of this analysis, the Tci2 genome

was not annotated. Tci2 contigs containing these RADmarkers/RADloci were manually annotated in Artemis v16.0.0 (McQuillan et al., 2011), using RNA-Seq data obtained from a NZ study (Choi et al., 2017), to identify genes up to 75 kb distant from the RADlocus. RNA-Seq fastq files were first checked using FASTQ Groomer v1.1.1 (Blankenberg et al., 2010) and subsequently aligned to the Tci2 genome using HISAT2 v2.1.0 (Kim et al., 2015). For this, the University of Glasgow Galaxy server was used (Afgan et al., 2018).

The translated nucleotide sequences of annotated genes were searched against the NCBI Nematode database, using BLASTP. Each alignment was considered on its own merit, however only predicted homologues which had an E-value of less than  $1e^{-5}$  were retained. All *C. elegans* homologues retained had greater than or equal to 20% pairwise identity, and *H. contortus* homologues had greater than or equal to 30% pairwise identity. All equivalent proteins retained from the *T. circumcincta* PRJNA72569 BLASTP had greater than 60% pairwise identity. In addition, BLAST tools in WormBase ParaSite were used to further identify *C. elegans* homologues, *H. contortus* homologues, and to identify equivalent genes in the PRJNA72569 *T. circumcincta* genome. If no gene could be identified in the PRJNA72569 *T. circumcincta* genome then the PRJNA72569 contig with the lowest E-value was recorded. Where relevant, wormbase.org was used to better understand the possible function and interactions of genes identified. Consensus fasta sequences for each individual (two sequences per RADlocus per individual) were extracted from the Stacks v2.0b output file and visually assessed in Geneious® v9.1.8 (<https://www.geneious.com>). Due to the time taken to manually annotate and identify genes, RADloci within the top 50 results of two or more statistical methods ( $F_{ST}$ ,  $F'_{ST}$  or  $D_{EST}$ ), were prioritised. In addition, the top five RADloci of each method were investigated. Lastly, the use of contig mean LD as an indicator for IVM selection was investigated by comparing contigs with SNPs of high  $F_{ST}$  by their mean  $F_{ST}$  and mean LD value. Results are described in sections 4.3.10 to 4.3.12.

## 2.8 Bioinformatics: Preliminary annotation of Tci2 genome

In January 2019, the Tci2 genome was annotated as part of this PhD. MTci2 RNA-Seq data generated from three biological replicates and two technical replicates

each of pooled adult females, pooled adult males and pooled sheathed L3 was provided by the BUG Consortium. Using STAR v2.5.2 (Dobin et al., 2012), RNA-Seq reads were aligned to the Tci2 genome. Following alignment, SAMtools and MultiQC were used to analyse the alignment success: between 87 and 90% of reads mapped successfully to the genome. The aligned reads and the Tci2 genome were provided as input to the program BRAKER v2.0 (Stanke et al., 2008; Hoff et al., 2016), composed of GeneMark-ET (Li et al., 2009; Barnett et al., 2011; Lomsadze et al., 2014), a program which annotates genes using aligned RNA-Seq reads, and AUGUSTUS (Altschul et al., 1990; Stanke et al., 2006; Camacho et al., 2009), which uses the pre-annotated genes to further predict additional coding sequences. Results are described in section 6.3.1. BUSCO v3.0.2 was used to assess the gene copy number (Simao et al., 2015). The annotated protein dataset (produced using the MTci2 RNA-Seq data) was analysed using -m proteins. The Tci2 genome assembly was also separately analysed using the nematoda\_odb9, *Caenorhabditis elegans*, as training material for AUGUSTUS. BUSCO results are provided and discussed in Chapter 7.

## 2.9 Bioinformatics: Pool-Seq

For the Pool-Seq experiment two farms were used. Farm 2 had a pre- and a post-treatment sample. Farm 3 had two technical replicates of the pre- and post-treatment samples. Following validation of the method, pairwise comparisons between the pre- and post-treatment samples were performed by farm to identify regions of genetic differentiation. Bioinformatics methods are detailed below. Note that scripts used are included in Appendix 2 when relevant, and are also available on GitHub (<https://github.com/SheepwormJM>). Raw data can be accessed via ENA using accession numbers ERS2487257 to ERS2487262.

### 2.9.1 Alignment of Pool-Seq reads to the reference genome

Fastqc and MultiQC were used to visually assess the quality of the sequencing data. Following this, reads were aligned to the Tci2 genome using the same script detailed in 2.7.2. For Pool-Seq, duplicate reads were marked. SAM files generated were merged and sorted, before conversion to BAM output files using SAMtools. Using the same method, reads were also aligned to the PRJNA72569

genome. MultiQC was used to assess the alignment success. Results are described in section 5.3.2.

## 2.9.2 Coverage

The genome-wide coverage depth was calculated as:

$$((\text{number of reads} \times 75 \text{ bp read length}) / \text{genome length})$$

for each sample population. In addition, SAMtools v1.3 *depth* was used to calculate the mean mapped depth per site, requiring a minimum mean read base quality (BQ) of 20 (-q20). To better appreciate the spread of coverage depth across the genome, SAMtools v1.3 *stats* was used to calculate the coverage frequency in bins of ten, ranging from 1 to 1000 read depth (-c(1, 1000, 10), i.e. the first bin included genomic sites with coverage 1-10, the next 11-20 etc.. Any sites with coverage depth of >1000 were binned together. Lastly, bedtools *makewindows* was used to split the genome into 6145 bp regions, based on the average size of a gene in the published reference genome PRJNA72569 (Quinlan and Hall, 2010; Quinlan, 2014; Choi et al., 2017). Following this, the read count per 6145 bp window was calculated by bedtools *multicov* (-f 37E-9 -q 20), requiring a mapping quality (MAPQ) of 20 or more, and that each read included in the analysis should overlap the region by at least 37 bp (equal to half the read length). Results are described in section 5.3.3.

## 2.9.3 Within population diversity and Tajima's D

An mpileup file which contained, for each genomic site, the coverage and sequencing quality of all mapped reads for each sample population was produced using SAMtools v1.3. Only reads with MAPQ > 20 and BQ > 30 were used. Duplicate reads were discarded. Each sample pileup was subsequently extracted using a custom script provided by Dr S. Doyle of the WSI. PoPoolation *Variance-sliding.pl* (Kofler et al., 2011a) was used to calculate the nucleotide diversity (Tajima's Pi) and Tajima's D. These tests are explained further alongside the results in section 5.3.4. Distinct 10 kb windows were used, requiring a minimum minor allele count of two, a minimum coverage of 10 and a maximum coverage of 400. The minimum covered fraction of each window was 0.1. A Wilcoxon sum

rank test, with continuity correction, was performed in R v3.1.2, to compare samples, and treatment groups (i.e. all pre- vs all post-treatment samples). Results are described in section 5.3.4.

#### 2.9.4 Pool-Seq genetic differentiation analysis

For this analysis the mpileup file (2.9.3) was used. Indels were identified using PoPoolation2 *identify-indel-regions.pl* (--indel-window 5, --min-count 2) (Kofler et al., 2011b). Basepairs adjacent to these were filtered (*filter-sync-by-gtf.pl*), to reduce the chance of calling false positive SNPs. After syncing the file (*mpileup2sync.jar*), PoPoolation2 *snp-frequency-diff.pl* was used to calculate major and minor allele frequencies for each variant site. The major allele frequencies were used to compare genetic relatedness of samples in R v3.1.2 using *cor()* and the package *Hmisc* to calculate Pearson's correlation coefficient (*r*). The packages *ggplot2* (Wickham, 2016) and *reshape2* (Wickham, 2007) were used to produce a correlogram using *geom\_tile*.

For each sample pairwise comparison (i.e. pre- vs post-treatment, within a farm) performed by PoPoolation2, variant site allele frequency differences were calculated and a Fisher's Exact Test was performed (*fisher-test.pl*). In addition, the degree of allele fixation between sample populations for each variant site was calculated ( $F_{ST}$  analysis of genetic differentiation, *fst-sliding.pl*), and data were smoothed. Distinct windows of 10 kb, 50 kb and 100 kb, allowing for fractional coverage of each window of 1, 0.5 and 0.1 were tested. Following this initial assessment, a window size of 10 kb and fractional coverage of 0.1 was subsequently chosen. Using the first 100,000 variant sites, allele frequency and coverage depth parameters for the  $F_{ST}$  analysis were trialled. Minimum minor allele counts of 2, 5 and 10 were compared. For all analyses ( $F_{ST}$ , allele frequency differences and Fisher's Exact Test), a minimum minor allele count of 5, (across all six samples/replicates), a minimum coverage of 10 (for each sample) and a maximum coverage equal to the top 2% of coverage for each sample was chosen (range 70 - 76x).

For all  $F_{ST}$  comparisons regions and variants of interest were defined as having  $F_{ST}$  values above a threshold cut-off of three standard deviations above the mean  $F_{ST}$  value for that comparison. Individual variant sites were considered of

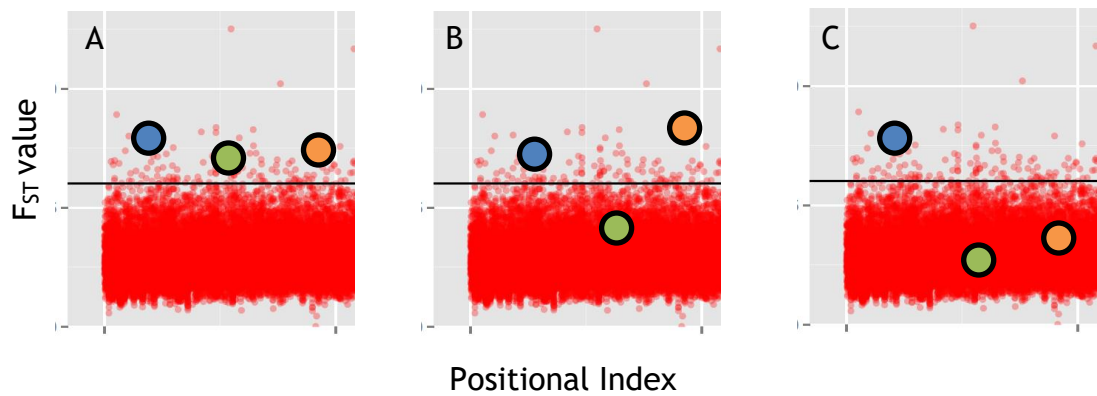


interest if they had a significant allele frequency difference ( $p < 0.05$ ), between the pre- and post-treatment samples. Data were always filtered within farm, rather than between farms. Contigs containing regions of interest were compared to contigs containing RADloci of interest from the ddRAD-Seq analysis. The  $F_{ST}$  analysis was repeated for reads aligned to the PRJNA72569 genome, and results were compared between genomes. In addition to within farm analysis, genetically differentiated regions identified in each farm analysis independently were compared using bedtools *intersect* to identify genomic regions of differentiation common to both farms. The expected number of 10 kb regions to be identified in both farm analyses by chance was calculated using:

$$\frac{(\text{Number of windows in Farm 2 data} \times \text{Number of windows in Farm 3 data})}{\text{Total number of 10 kb windows (68,500)}}$$

The expected number of windows identified on both farms was compared to the observed number of windows identified on both farms using a Chi-squared test in Excel.

Farm 3 had two technical replicates for each sample, referred to as Pre-IVM Replicate A and Replicate B, and Post-IVM Replicate A and Replicate B. These were used to additionally filter false positive results (see Figure 2.3). Each pre-IVM replicate was compared to each post-IVM replicate. Windows and SNPs were filtered so that only results identified as being of interest in all of these four pairwise comparisons (i.e. Pre-A vs Post-A, Pre-A vs Post-B, Pre-B vs Post-A and Pre-B vs Post-B), were retained. In addition, the two post-IVM technical replicates on Farm 3 were compared for regions of genetic differentiation. It was assumed that any region identified would not be related to IVM treatment and these regions were discarded from further Farm 3 analyses. Results are described in section 5.3.5 and are compared with ddRAD-Seq results in 5.3.6.



**Figure 2.3: Schematic representation of filtering of Farm 3 data. Shown are three Manhattan plots. The black line indicates the threshold calculated for the pairwise comparison plotted. (A). Pre-A vs Post-A, (B). Pre-A vs Post-B, (C). Post-A vs Post-B. Three 10 kb windows are indicated with  $F_{ST}$  values greater than the threshold for the pre- to post-IVM comparison in plot (A). The blue dot represents a 10 kb window which is above the threshold in both (A) and (B), but also in (C), suggesting that it is related to inherent diversity within the sample population rather than indicating IVM selection. It is discarded. The green dot only appears to be selected in (A), and not in (B), so it too is discarded. Lastly, the orange dot is high in both (A) and (B), but not in (C). It is retained for further analysis.**

In addition to the above, a Cochran-Mantel-Haenszel (CMH) test was performed (PoPoolation 2, *cmh-test.pl*) to identify regions significantly ( $p < 0.0001$ ) and consistently differentiated by allele frequency on both farms. This was performed five times, each time comparing Farm 2 (Pre- vs Post-IVM) to a different Farm 3 technical replicate comparison (Pre- vs Post-IVM or Post- vs Post-IVM). As for the  $F_{ST}$  analysis, results were filtered for consistency across all Farm 3 replicate comparisons. Results were excluded when identified using the Farm 3 post-post technical replicates, which should not have differences related to IVM selection. Results are described in section 5.3.5.

### 2.9.5 Identifying genes within genetically differentiated regions

Genetically differentiated regions identified on a farm were further examined. The amino acid sequences of genes annotated by AUGUSTUS within, or crossing a window of interest were extracted using bedtools *intersect*. Homologues (*C. elegans* and *H. contortus*) and equivalent proteins (proteins identified within the PRJNA72569 *T. circumcincta* genome) were identified using the WSI *farm\_blast* BLASTP. As hundreds of proteins were investigated, only the top hit was retained (blast\_options "-max\_target\_seqs 1"). In some instances this may have permitted selection of a shorter alignment, with a lower (and hence better) E-value, than a more appropriate, longer alignment. However, in most cases this method was

expected to be successful. Results were further filtered to retain only proteins with E-value  $< 1e^{-05}$ . Homologues for *C. elegans* were accepted if they had 20% pairwise identity or more, or for *H. contortus* if they had 30% pairwise identity or more. No equivalent proteins identified in the *T. circumcincta* PRJNA72569 genome were retained with less than 50% pairwise identity, however most had greater than 80% identity. Predicted homologues or equivalent proteins which were common to both farms were identified. All results are provided in Appendices 5, 6 and 7 and are described in sections 6.3.2 and 6.3.4.

### 2.9.6 Obtaining gene ontology information for identified genes

Following identification of homologues in *C. elegans* and *H. contortus* and equivalent proteins in PRJNA72569 *T. circumcincta*, BIOMART (Smedley et al., 2015) as hosted by WormBase ParaSite, was used to identify annotated gene ontology (GO) terms. GO terms were searched for in a hierarchical manner to attempt to obtain the maximum information possible, as many genes in the PRJNA72569 *T. circumcincta* genome are missing protein information.

To do this, each list of proteins previously identified using BLASTP (2.9.5), was provided to BIOMART, beginning with the PRJNA72569 equivalent protein list. In addition to GO terms, homologues in *H. contortus* and *C. elegans* and the gene name were included in the output file. From this file, *H. contortus* homologues identified, with no *C. elegans* homologue, were subsequently included when the *H. contortus* protein list was provided to BIOMART. For this list, the *C. elegans* homologue was requested in addition to the GO terms and gene name. Any *C. elegans* homologues identified by BIOMART from either search were included in addition to the *C. elegans* homologue list previously identified (2.9.5) and were provided to BIOMART to identify GO terms. All results are provided in Appendices 8 and 9. Genes of interest were then identified using the following search terms; Neuronal genes ('neuro', 'nerve', 'axon', 'dendrite', 'synap' and 'channel'), Lipid metabolism genes ('lipid', 'ligase', 'glycerol', 'glycerone', 'fatty', 'lipo' and 'cholesterol'), Male ('male' and 'sperm'), Female ('female', 'egg', 'vulva' and 'zygot'). Genes were manually checked. Results are provided in section 6.3.2.

### 2.9.7 Identifying SNPs of high or moderate impact

The program SnpEff (Cingolani et al., 2012), was used to assess the possible impact of SNPs within high  $F_{ST}$  10 kb windows. First, a *T. circumcincta* database was created using SnpEff following which SnpEff was run for each farm separately, using default values. The program requires both gene annotations and a list of variant sites in VCF format. SAMtools v1.3 was used to produce a VCF file of all variant sites with MAPQ > 20 and BQ > 30. Mean depth per variant site was determined for each farm separately using VCFtools v0.1.14. The VCF file was filtered to retain only variant sites with coverage depth of between 20 to 60x inclusive, as determined for each farm independently. These variant sites were used as input for SnpEff. The SnpEff annotated VCF output file was filtered using bedtools *intersect* to retain only SNPs with: (i) high  $F_{ST}$ , (ii) a significant allele frequency difference ( $p < 0.05$ ) and (iii) which were within the high  $F_{ST}$  10 kb windows. Lastly, each file was filtered using *grep* to identify those SNPs which were deemed to be of 'HIGH' or 'MODERATE' impact. Results are provided in section 6.3.3.

### 2.9.8 Comparison with other GWAS

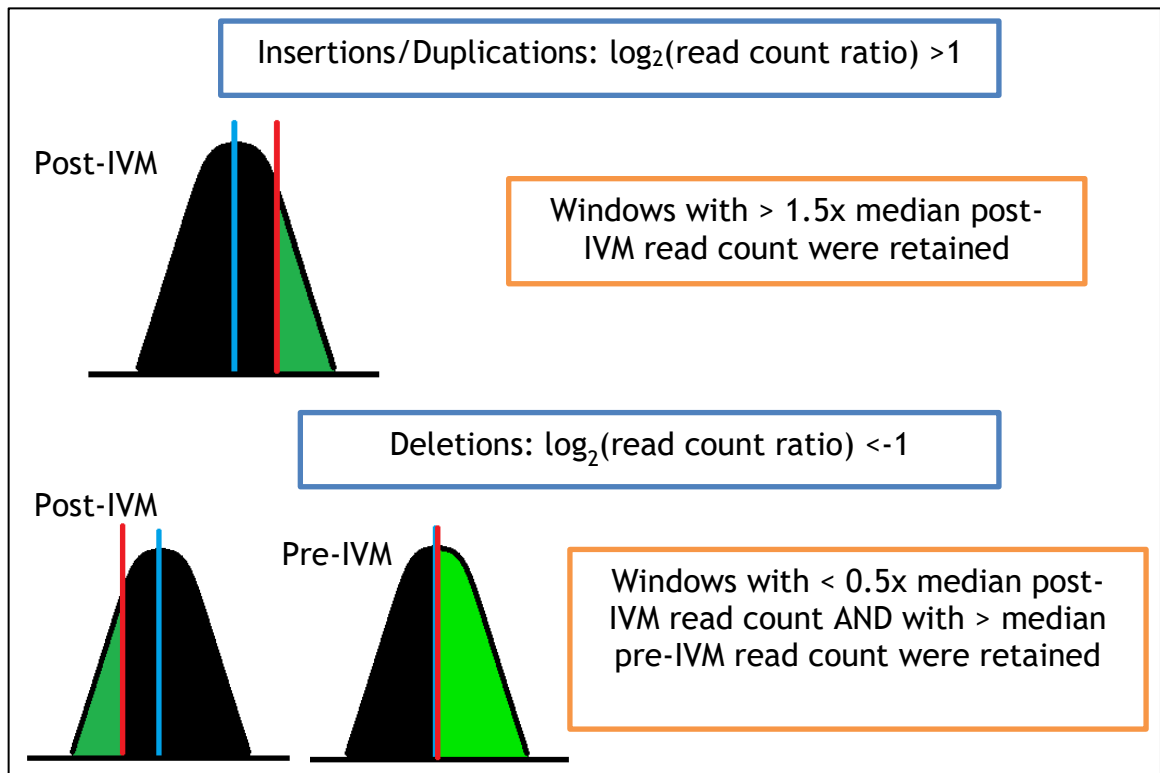
Three recently published GWAS investigating IVM resistance in *T. circumcincta* (Choi et al., 2017), *H. contortus* (Sallé et al., 2019), and *O. volvulus* (Doyle et al., 2017) were compared to results obtained in the present study. In addition, recent work using a genetic cross of MHco3(ISE) and MHco18(UGA) to investigate IVM resistance in *H. contortus* was also compared (BUG Consortium). To compare between studies, the genes, SNPs or regions of interest, as identified by the study authors were compared with homologues of Tci2 genes identified within regions of genetic differentiation in the present study. Results are described in section 6.3.5. All results are presented in Appendices 12-14.

### 2.9.9 Copy number variation

Variation in copy number between sample populations was assessed using a proxy method dependent on the read count within a given window size (Xie and Tammi, 2009). Prior to performing this analysis a file containing only reads which had successfully aligned to the Tci2 genome was produced using SAMtools *view* (-F4). The resulting file was piped to a perl script to produce a 'hits' file as

required by the program CNV-seq (Xie and Tammi, 2009). Copy number variation (CNV) was calculated using CNV-Seq *cnv-seq.pl* specifying a genome size of 680 Mb and requiring a set window size of 5 kb. Each window overlapped the previous window by 2500 bp. For each window, the  $\log_2$ (read count ratio) between the pre- and post-IVM samples was calculated. Results were filtered to retain only windows with  $\log_2$ (read count ratio) greater than one or less than minus one, which were significant at  $p < 0.00001$ . Windows that had no reads in one sample of a comparison were discarded. For each comparative analysis, the post-IVM sample was compared to the pre-IVM sample. Hence, if the  $\log_2$ (read count ratio) is positive, read counts are higher in the post-IVM sample compared to the pre-IVM sample. A two tailed, paired t-test was performed in Excel to compare the pre- and post-treatment sample means.

Based on the genome-wide average read count per window, the read count distribution and the median read count, the results were then further filtered as described in Figure 2.4. As CNV is less accurately determined using Pool-Seq data than for individual sequencing, filtering aimed to retain only CNV of large magnitude. For insertions/duplications ( $\log_2$ (read count ratio)  $> 1$ ), windows were only retained if the post-IVM sample had a read count greater than 1.5x the median post-IVM sample read count. For deletions, windows were only retained if they had read counts of less than half the median read count in the post-IVM treatment samples and had a read count greater than or equal to the median read count in the pre-IVM treatment sample. For Farm 3, CNV were also required to be present in all four pre- to post-IVM replicate comparisons. *Tci2* genes were identified for each window retained after filtering using *bedtools intersect*. Homologues and equivalent proteins were identified using the *WSI farm\_blast* BLASTP tool in the PRJNA72569, *H. contortus* and *C. elegans* protein databases.



**Figure 2.4: Schematic diagram showing filtering of CNV regions. The green filled areas indicate the retained read-counts. Results were filtered as follows: for insertions/duplications (top), any windows with read counts  $> 1.5x$  post-IVM sample median read count were retained. For deletions (bottom), windows retained were required to have a read count of  $< 0.5x$  post-IVM sample median read count AND the median read count (or greater) of the pre-IVM sample median read count. The blue line indicates the median read count, the red line the cut-off.**

Windows retained were considered for potential relevance and robustness by a number of factors. The contigs and their corresponding gff3 data were extracted for visualisation with aligned reads in Artemis. Four points were then considered. Firstly, the length of the CNV was determined (although the sliding windows used were 5000 bp, the read count across the window was not necessarily uniform). Secondly, the MAPQ score of the reads was visually assessed to obtain an estimate of mapping quality over the region. Thirdly, the position of the CNV relative to gene annotations was considered (i.e. was it within an intron, exon or close to an annotated gene). Lastly, was the CNV region within, or covered by, a repetitive or low-complexity region by more than 70%, which might indicate falsely aligned reads (Troost et al., 2018)?

To identify CNV of specific P-gp genes, a BLASTP search, with no restrictions, using the WSI *farm\_blast* was performed of all translated nucleotide sequences of genes annotated by AUGUSTUS. Using PRJNA72569 equivalent proteins, and C.

*elegans* protein homologues of P-gps, relevant results were extracted and manually examined to identify all Tci2 P-gps required. Where necessary, amino acid sequences were extracted and aligned in Geneious® v9.1.8, and phylogenetic trees were plotted. CNV read counts were subsequently extracted for the appropriate regions. Results are described in section 6.3.6.

### 3 Don't put all your eggs in one basket: The role of diversity in PGE and anthelmintic resistance on a commercial farm

#### 3.1 Abstract

Commercial sheep flocks vary widely in their size, management and purpose. However parasitic gastroenteritis is common to all. This disease is complex with multiple strongyle species potentially contributing to clinical signs. Various management and breeding practices are known to affect the clinical outcome of infection. Almost all farmers will use anthelmintics to control and treat the disease, but efficacy can decline with use. Despite this, few farmers routinely test the anthelmintic efficacy of the products they use. In addition, multiple factors can affect the interpretation of results of such tests. I used three different classes of tests (*in vivo* - faecal egg count reduction test, *in vitro* - egg hatch test and larval development test, and molecular test - pyrosequencing) to investigate anthelmintic efficacy on a typical commercial farm in the UK. In addition, parasite abundance and species composition was assessed over time to gain insight into how such tests might be affected over the farming year. Lastly, microsatellites were used to assess for evidence of selection or change in the *T. circumcincta* population present on the farm over time. My results showed that despite modern farming practices, a considerable diversity of strongyle species can be maintained. Both the *in vivo* and *in vitro* tests of anthelmintic efficacy were considerably confounded by this species diversity. By using a combination of tests it was identified that *T. circumcincta* individuals were highly dual-resistant to both BZ and IVM anthelmintics. Genotyped offspring of individuals surviving BZ treatment indicated survival of heterozygous resistant individuals. No change over time in the *T. circumcincta* individuals was noted using microsatellites, and individuals were as diverse as had been found on other farms in previous studies. Neither the *in vivo* nor *in vitro* tests used to detect IVM resistance were sensitive to the presence of a highly resistant species within the parasite community. A genetic test for IVM resistance, as is currently available for BZ resistance, would therefore greatly help research into IVM resistance.



## 3.2 Introduction

Parasitic gastroenteritis (PGE) is a significant production and welfare concern for sheep farmers in the UK (Sargison, 2014; APHA, 2018). Various means of disseminating advice to farmers on managing and monitoring PGE and anthelmintic efficacy are available. The best developed resources are provided by 'SCOPS' (Sustainable Control of Parasites in Sheep), including a website ([www.scops.org.uk](http://www.scops.org.uk)), best practice guidelines for farmers, and a technical manual for veterinary surgeons (Abbott et al., 2012). Uptake of these recommendations varies considerably (McMahon et al., 2013c; Learmount et al., 2016b; Jack et al., 2017) and anthelmintic resistance can still arise even with their use (Learmount et al., 2016a).

Farmers monitor PGE in a variety of ways (McMahon et al., 2013c). These may include noting clinical signs of illness within the flock, monitoring weight gain over time, obtaining faecal egg counts (FECs) or receiving post-mortem reports. In practice, a combination of these methods is most effective in fully understanding the presence and impact of PGE on a farm. For example, FECs do not always directly correlate with intensity of infection. This may be because infection is not yet patent, or because there is variation between species, duration of infection or the host immune response (Coyne et al., 1991; Stear and Bishop, 1999; Saccareau et al., 2017). In the case of *T. circumcincta*, the density of infection affects the fecundity of adult females (Bishop and Stear, 2000). Similarly, low weight gain can be caused by many nutritional and disease-related factors.

Following collection of a faecal sample, eggs are harvested and 'strongyle eggs' are counted. Strongyle eggs describe those of a similar appearance but encompass many different species of Trichostrongylid parasites of ruminants. There are five FEC methods in use with varying sensitivities: the McMaster (15 to 50 eggs per gram (epg)) (Coles et al., 1992; Cringoli et al., 2004), FLOTAC (1 epg) and Mini-FLOTAC (5 to 10 epg) (Rinaldi et al., 2014; Godber et al., 2015), FECPAK (30 epg) (Coles, 2003) and a cuvette method (1 epg, Christie and Jackson (1982), see Chapter 2.2.2). The Mini-FLOTAC and FECPAK are commercialised for use on farm. Neither require a centrifuge and the FECPAK can be performed without a microscope; instead an image is taken allowing

interpretation by an expert distant to the farm, making it highly practical for use by farmers. Dependent on the reason for sampling, each of these methods will vary in their applicability. For example, to monitor for the presence of high FECs requiring treatment the McMaster or FECPAK is suitable, however, to determine anthelmintic efficacy, it is better to use a test with high sensitivity.

The current gold standard test to assess anthelmintic efficacy, a 'controlled efficacy test', requires sacrificing animals (Coles et al., 2006) so is rarely used. The faecal egg count reduction test (FECRT), described in Figure 3.1, is simple to perform and is the most frequently used method on farm, but it can be expensive and labour intensive, as it requires setting aside groups of lambs for use in the test. The meat (or milk) withdrawal times of the anthelmintics used in the FECRT must also be observed. A modification of the full FECRT is to perform a 'drench test', sampling a group of animals post-treatment only, although interpretation is limited even with the inclusion of an untreated control group (Coles et al., 2006). For both the FECRT and drench test, analysis should ideally be done using separate FECs for each individual, obtaining an arithmetic mean (Dash et al., 1988). Although a simple calculation can be used as shown, with variations on this (McKenna, 2006), it is also possible to calculate the FECR using zero-inflated Bayesian methods, which can be more robust to inherent variation and bias (Wang and Paul, 2017; Wang et al., 2017). It is important to realise the limitations in determining anthelmintic efficacy from a FECRT. In particular the possibility of worms remaining post-treatment, but which shed fewer eggs compared with before (Sutherland et al., 2003a), or the presence of anthelmintic resistance in only one or two particular strongyle species within the population (McKenna, 1997).

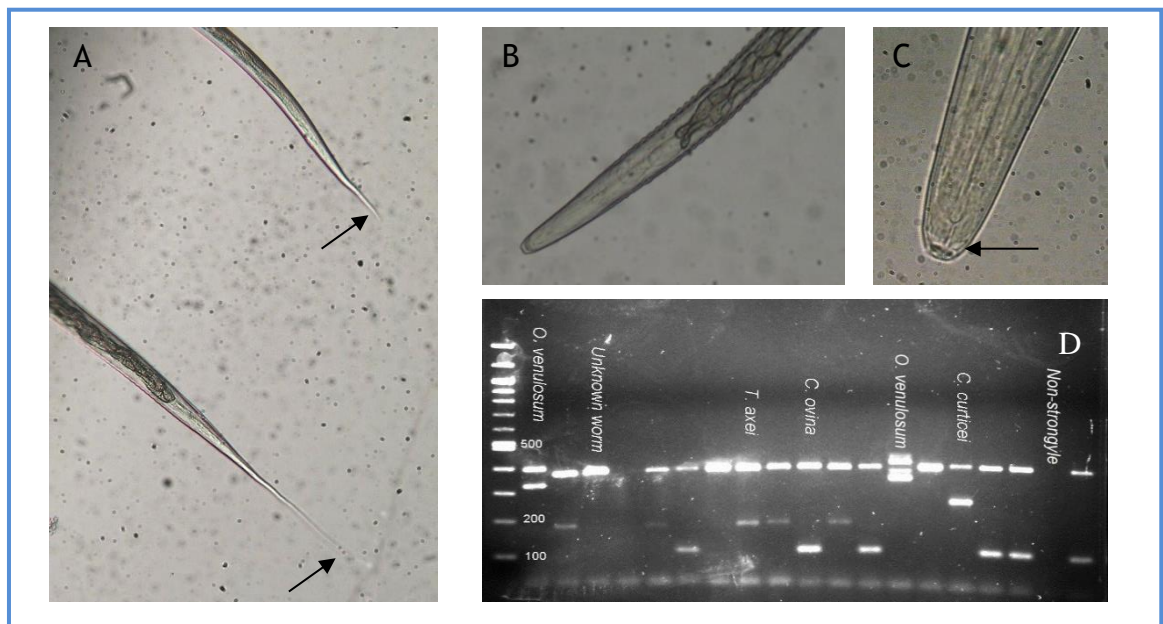
FECRT guidelines:

1. Identify groups of 10 or more sheep for each anthelmintic used +/- a control group.<sup>1</sup>
2. Calibrate equipment (weigh scales and dosing gun)
3. Weigh each animal individually.
4. Collect faecal sample, ideally *per rectum*, of roughly 3-5 g weight.
5. Dose each animal to their individual weight.
6. Perform FEC.
7. Re-sample animals 7 to 17 days later depending on the anthelmintic used:
  - 1-BZ: 10 - 14 days post-treatment
  - 2-LEV: 7 days post-treatment
  - 3-ML: 14 - 17 days post-treatment
  - 4-AD: 14 days post-treatment
  - 5-SI: 14 days post-treatment<sup>2</sup>
8. Perform FEC and calculate difference. Various equations may be used including:

$$\text{FECR \%} = ((\text{FEC pre-treatment} - \text{FEC post-treatment}) / (\text{FEC pre-treatment})) * 100$$

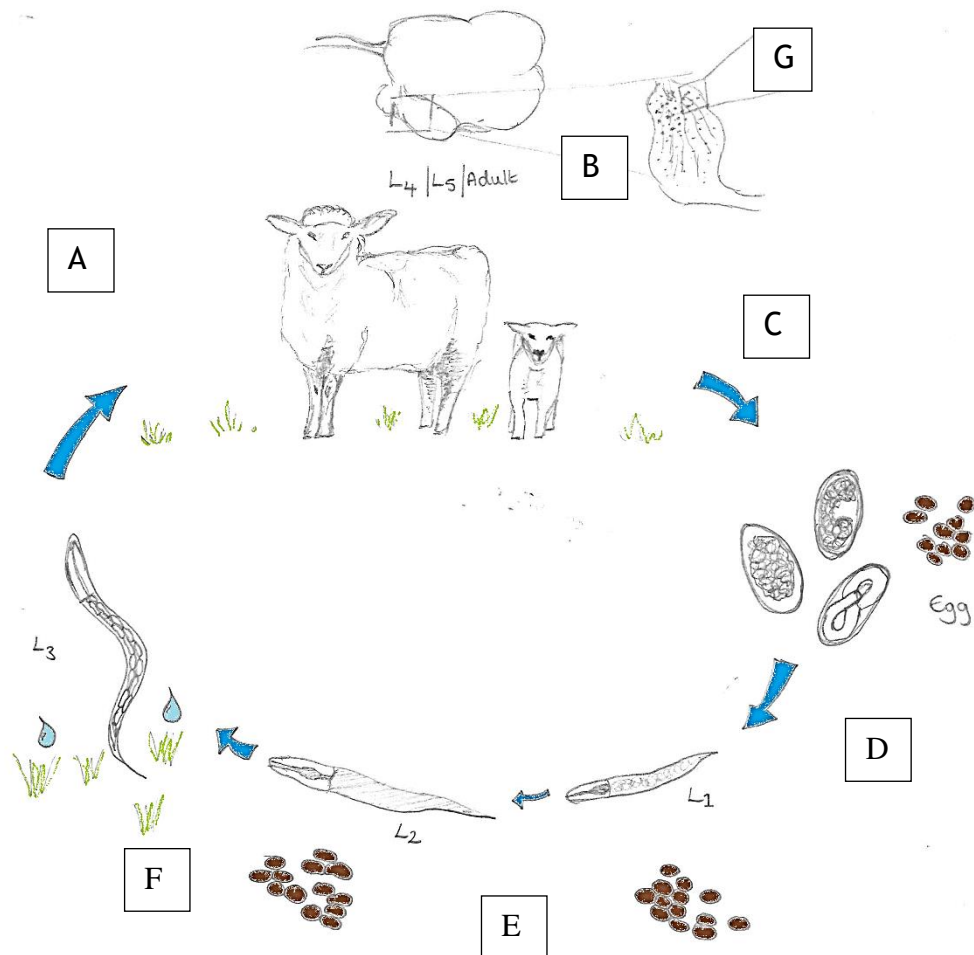
**Figure 3.1: Faecal egg count reduction test guidelines adapted from Coles et al. (2006), with the addition of the 4-AD (Kaminsky et al., 2008) and 5-SI anthelmintic classes (Little et al., 2011). <sup>1</sup>If using animals older than six months of age it is important to check that the strongyle faecal egg count is > 150 epg. <sup>2</sup>Note that not all L4 larvae of every species may be killed.**

Strongyle species identification can be performed using morphology of the L3 larval stages (Crilly and Sargison, 2015). L3 are killed and examined under a microscope, in particular focusing on sheath tail length and shape, head shape, the presence of refractile bodies within the head, the number of gut cells and the overall length of the larva (Figure 3.2 A-C) (van Wyk and Mayhew, 2013). This can be an excellent means of identification by skilled parasitologists, but is time consuming and requires training. Molecular methods have been developed, using the internal transcribed spacer region (ITS2) of the ribosomal RNA gene. Worms can be identified by individual species PCR (Wimmer et al., 2004), multiplex PCR (Bisset et al., 2014) (Figure 3.2 D) and multiplex-tandem PCR, commercialised as AusDiagnostics (Roeber et al., 2012; Roeber et al., 2017) using eggs or larvae as template. Newer methods use next generation sequencing (NGS) to sample the entire 'nemabiome' using strongyle specific, but species generic primers which amplify the ITS2 region, prior to pooled sequencing and extrapolation of the approximate species proportions by read count (Avramenko et al. (2015), [www.nemabiome.ca](http://www.nemabiome.ca)).



**Figure 3.2: Strongyle species identification methods. Morphological identification of L3: A.** By sheath tail length, e.g. top = short tailed, *Teladorsagia circumcincta* type, bottom = long tailed, for example *Chabertia ovina*, **B.** By head shape, i.e. rounded, square, bullet shaped, **C.** *Cooperia curticei* head demonstrating the presence of refractile bodies (arrow). **Molecular methods: D.** Image of PCR products from a multiplex PCR reaction. The top band is the amplicon of the generic primers, while the lower band is species specific (Bisset et al., 2014).

The presence of multiple species contributing to PGE complicates both management and research on farm. Typically, parasites reach patency within three weeks following ingestion, but there is considerable variation between species (Roerber et al., 2013b), time of year (Gibbs, 1986a), and immune response of the host (Smith et al., 1985). Free-living stages pass through three developmental stages (egg, L1 and L2) before reaching the infective L3 stage (Figure 3.3). The development period and survival time of L3 on pasture is species specific, but is also affected by relative humidity, moisture and soil temperature (Gibson and Everett, 1972; O'Connor et al., 2006; van Dijk and Morgan, 2011; Wang et al., 2014). In the case of *T. circumcincta*, L3 were able to survive on pasture for 16 months from the time of the egg being laid (Gibson and Everett, 1972), and L3 can be stored at 8 °C for longer than this.



**Figure 3.3: *Teladorsagia circumcincta* lifecycle as an example of a typical Trichostrongylid lifecycle. A. An infective L3 on herbage is ingested. B. The L3 exsheaths within the rumen (Bekelaar et al., 2018) and enters the gastric glands of the abomasum (Sommerville, 1953). Here it moults twice and emerges as an L5 before reaching patency as an adult. C. Following patency (typically estimated to take 17-21 days, although 14 days has been observed for *T. circumcincta*), eggs are shed into the faeces (Barrett et al., 1998). D. *T. circumcincta* eggs develop at temperatures above 4 °C (Crofton, 1965). E. The L1 larvae hatch and moult to L2, usually within the space of a few days. F. Following climatic cues, the infective L3 emerges from the faecal pat onto herbage (Gibson and Everett, 1972). Here it may move up the sward, or down into the soil dependent on local environmental conditions (van Dijk and Morgan, 2011). G. Close up of an opened abomasum, illustrating the mucosal folds (lines), with the multi-papillary raised lesions associated with an infection of *T. circumcincta*. These lesions are due to the presence of maturing L4 within the gastric glands, predominantly within the fundic region. Not all L4 will mature – some will be suppressed by the host's immune response, others will become hypobiotic overwinter and re-emerge in spring.**

The determination of anthelmintic efficacy by FECRT on a farm may vary across the year dependent on which species are sensitive, as seasonal variation in species prevalence is recognised on farms (Boag and Thomas, 1977). In addition, the proportion of sensitive to resistant nematodes within a host at any given time will also affect the results of a FECRT. How readily resistant parasites are 'diluted' within hosts is largely dependent on the number of sensitive parasites

*in refugia* on farm, which the host is likely to ingest between treatments throughout the year (van Wyk, 2001). Monitoring anthelmintic resistance on a farm over time to determine treatment choice, and to understand how *refugia*, climate and management practices affect the prevalence of resistance is thus complicated. FECRTs are expensive, in terms of actual costs and labour time, but also unavoidably apply drug-selection to a subset of animals on a farm. Various *in vitro* bioassays are available to measure anthelmintic resistance (Gill et al., 1995; von Samson-Himmelstjerna et al., 2009a) (see Introduction, Table 1.2), which avoid drug selection on farm. Unfortunately, these are also expensive, time consuming, can be technically challenging, and in most cases, are not commercially available.

To overcome current limitations in resistance diagnostics there is a pressing need for sensitive and specific molecular tests, but these require knowledge of the genetic loci conferring resistance (e.g. BZ SNPs (Kwa et al., 1995; Skuce et al., 2010)). In order to develop such tests it is also important to understand the underlying parasite population structure, its diversity, and how this changes with time e.g. under different climatic conditions and management strategies. A well-established method to monitor changes in a parasite population over time is the use of microsatellite markers. Microsatellites are short, rapidly evolving DNA sequences of tandem repeat units, most commonly di- or tri-nucleotides. Thirteen microsatellites have been characterised in *T. circumcincta* and used to measure within and between-population diversity on UK farms (Grillo et al., 2006; Grillo et al., 2007; Redman et al., 2015).

The pattern of PGE and the involvement of different species over time were monitored on a commercial farm in Scotland. Changes in BZ resistance markers and neutral microsatellite markers in *T. circumcincta* were measured and compared with results from FECRTs and bioassays to determine the ease and effectiveness of resistance diagnosis by commonly used methods.

## 3.3 Results

### 3.3.1 Management, parasite abundance and diversity over a year on a farm

#### 3.3.1.1 Study farm and flock management

Farm 1 was a commercial, 150 acre lowland farm in southeast Scotland with previously diagnosed anthelmintic resistance. The farm was studied from March 2016 to March 2017. The breeding flock consisted of approximately 370 Cheviot Mule ewes which were bred to Texel, Suffolk and Beltex rams. Lambs were finished on pasture before being sold at market in the autumn. Lambing began in late March, with ewes housed on straw from early January, and continued for approximately five to six weeks, with an average annual lambing percentage of 180% (Table 3.1). Ewes were fed haylage (analysed for nutritional content by the farmer) and supplemented with sheep nuts (Davidsons; Super Ewe Rolls (20% CP) and EWELAC 18 Nuts (18% CP)) whilst indoors, and sheep nuts continued to be provided for a short period following lambing. Ewes and new lambs were returned to pasture within 48 hours of lambing.

Pastures were grazed intensively by the flock, with frequent rotation of the small group of rams between pastures. Ewes and lambs were kept together in large fields, with little to no rotation until weaning in late August, when all lambs were re-grouped as a single flock and moved to a 'clean' pasture. This had been previously used for silage, but had been occasionally grazed by rams during silage growth and gimmers (young adult females which have not reproduced) had grazed the pasture earlier in the year. Ewes were re-sorted onto several pastures and in the autumn were mixed with the rams for breeding. New rams were bought in September 2016 and replacement ewe lambs were obtained in October 2016. The management observed on the farm over the year of study (Table 3.1) was similar to that in previous years. Occasionally, in previous years, when grass sward height was sufficient, pastures on the farm had been rented to neighbouring cattle farmers. Adult females were vaccinated with Covexin 10 (Zoetis) in February/March 2016 and with Enzovax (MSD Animal Health) and Toxovax (MSD Animal Health) in the summer (Table 3.1). These vaccines protect against infection with clostridia, *Chlamydophila abortus* and *Toxoplasma gondii*,

respectively. Sheep were generally healthy although lameness was a concern - this was primarily noticed in the rams.

**Table 3.1: Main management events on Farm 1.**

Month	Management event
January	Ewes housed
February	Ewes housed. Pregnant sheep vaccinated with Covexin 10.
March	Started lambing. Ewes put outside to pasture when lambed.
April	Finished lambing
May	
June	Lambs treated with Vetrazin (Elanco) to protect against blowflies.
July	
August	Weaned lambs as a single group onto silage aftermath. Ewes placed in several groups. Gimmers vaccinated with Enzovax and Toxovax.
September	New rams bought in.
October	Replacement ewe lambs bought in. Breeding began.
November	Last few lambs sold. Breeding continued.
December	

### 3.3.1.2 Study farm anthelmintic treatments

Anthelmintic dose was chosen by the farmer following weighing of the sheep estimated to be the heaviest in the group, and all sheep were subsequently dosed to this weight then returned to the same pasture. Dosing guns were rarely calibrated. The shepherd would decide to dose sheep based on clinical signs of scour and/or concern of potential PGE (possibly including a FEC) and would usually choose the anthelmintic product based on veterinary recommendation.

The majority of the ewes were treated following parturition with MOX (Cydectin 0.1%; persistent action against reinfection by susceptible *T. circumcincta*, *H. contortus* and *T. colubriformis* for five weeks, Zoetis), before being returned to pasture (Table 3.2). The shepherd aimed to treat all ewes with triplets, and some ewes with twins, based on assessment of need (e.g. body condition score), but not ewes with singles. However, in practice, treatments were more widespread and the medicine book recorded that 350 (out of 370) ewes were treated with MOX. Samples collected from ewes and lambs between lambing and weaning were from a group which included ewes dosed with MOX. Rams were treated in February using LEV, in July with IVM and following the breeding period in late autumn (anthelmintic unknown) (Table 3.2).



Following identification of *Nematodirus battus* eggs in the faeces of four young lambs in late May, lambs were treated with albendazole three times in May, June and July, at each time point leaving 5-10% of lambs untreated (Table 3.2). Although no FECRT was performed at these treatment times, FECs were conducted within 11 to 15 days following each treatment and at each time point strongyle eggs were detected. In late September, lambs were dosed with LEV by the farmer. In addition, FECRTs were assessed in a small group of 32 lambs at this time point, to assess efficacies of albendazole (Albex 2.5%, Chanelle Vet UK), LEV (Levacide 3%, Norbrook) and IVM (Noromectin 0.08%, Norbrook), as part of this study (see 2.2.5). Replacement ewe lambs and new rams were treated with MPTL and doramectin on arrival at the farm. The ewe lambs were kept on pasture for a fortnight before being moved to rented grazing. The new rams were integrated with other rams after 48 hours. MPTL has been used for quarantine treatment of new stock on this farm since at least 2013. The medicine record book indicates that in 2013 and 2015 replacement ewe lambs were dosed with MOX rather than doramectin. In 2012, LEV was used for quarantine treatment of ewe lambs.

**Table 3.2: Anthelmintic treatments during 2016 on Farm 1.**

Date	Anthelmintic (Product)	Sheep group	FECRT efficacy (95% HPD interval) <sup>1</sup>
04/02/2016	LEV (Levacide 3%)	Rams	ND
01/04/2016 to 28/04/2016	MOX (Cydectin Oral)	Lambled ewes - primarily triplets, but also some twin/single bearing ewes	99.6% (99.5-99.7%)
23/05/2016	BZ <sup>2</sup> (Albenil)	Lambs (excluding 5-10% of lambs)	ND
14/06/2016 and 15/06/2016	BZ (Albenil)	Lambs (excluding 5-10% of lambs)	ND
01/07/2016	IVM (Noromectin)	Rams	ND
06/07/2016 and 07/07/2016	BZ (Albenil)	Lambs (excluding 5-10% of lambs)	ND
09/09/2016	MPTL (Zolvix) Doramectin (Dectomax)	Bought in rams	ND
20/09/2016	LEV (Levacide 3%)	Lambs	ND
20/09/2016	BZ (Albex 2.5%)	Ten lambs	65.3% (14.5-86.2%)
20/09/2016	LEV (Levacide 3%)	Ten lambs	80.8% (58.0-91.1%)
20/09/2016	IVM (Noromectin 0.08%)	Twelve lambs	77.2% (45.5-91.6%)
07/10/2016	MPTL (Zolvix) Doramectin (Dectomax)	Ewe lambs	ND

All anthelmintic treatments given to the flock are shown. The product name, as obtained from the farm medicine record book, is also provided. <sup>1</sup>For Faecal egg count reduction tests performed during the study the anthelmintic product name and efficacy, as calculated using 'eggCounts' shiny is provided with the 95% Highest Posterior Density interval in brackets (Torgerson et al., 2014; Wang and Paul, 2017; Wang et al., 2017). <sup>2</sup>BZ = Albendazole. ND = not done.

### 3.3.1.3 Faecal egg counts over the year

Individual FECs were conducted on approximately ten ewes every three weeks throughout the year from 20<sup>th</sup> March 2016. In addition, lambs were sampled every three weeks from 16<sup>th</sup> May 2016 to 23<sup>rd</sup> November 2016 when the last lambs were sold. Again, ten samples were collected at each time point excluding the first, when only four lambs were sampled. Arithmetic means were calculated. From mid-April until weaning the same pasture group of ewes and lambs were sampled. The aim was to sample the group generally rather than to follow ten individuals across the year. This was partly due to practical

limitations of working with a commercial flock, but also to assess the flock as a whole and reduce host bias. Following weaning, new groups of ewes and lambs were followed. As much as possible, these new groups were consistently followed from this point forward until housing, when a new group of ewes was identified and followed until lambing. The occasional sample was collected by K. Hamer or Prof. N. Sargison. Results of the FECs are included in Hamer et al. (2019).

Overall, FECs were comparatively low, remaining below 200 epg for much of the year in both ewes and lambs. Ewe FECs rose from a mean of 452.1 epg on the 20<sup>th</sup> March, to 912.0 epg three weeks later when a further ten ewes were sampled shortly after parturition (Table 3.3, Figure 3.4). The increase was not statistically significant, likely reflecting the wide range of individual counts - the median FECs varied less, being 409.5 and 675.0 epg respectively. A MOX FECRT was performed on the 7<sup>th</sup> April 2016 by the farm veterinary surgeons. The mean strongyle FEC decreased from 912.0 epg to only 3.4 epg, an efficacy of 99.6% (95% Highest Posterior Density (HPD) Interval: 99.5 to 99.7%) (Torgerson et al., 2014; Wang and Paul, 2017; Wang et al., 2017). Individual reduction percentages ranged from 98.9 to 100% reduction (Table 3.2). Further ewe mean FECs remained low within the pasture group followed, ranging between mean 2.8 epg and 129.8 epg (Figure 3.4). This latter value was from a group sampled on the 7<sup>th</sup> June 2016 and was higher compared with other samples collected throughout the summer and early autumn. A rise in egg output from ewes during the autumn breeding period was also observed, rising from a mean 23.0 epg on the 10<sup>th</sup> October to 106.2 epg on the 23<sup>rd</sup> November (Figure 3.4). FECs subsequently reduced, such that in January, February and early March 2017 FECs were low, prior to rising again in late March at lambing time (Figure 3.4).

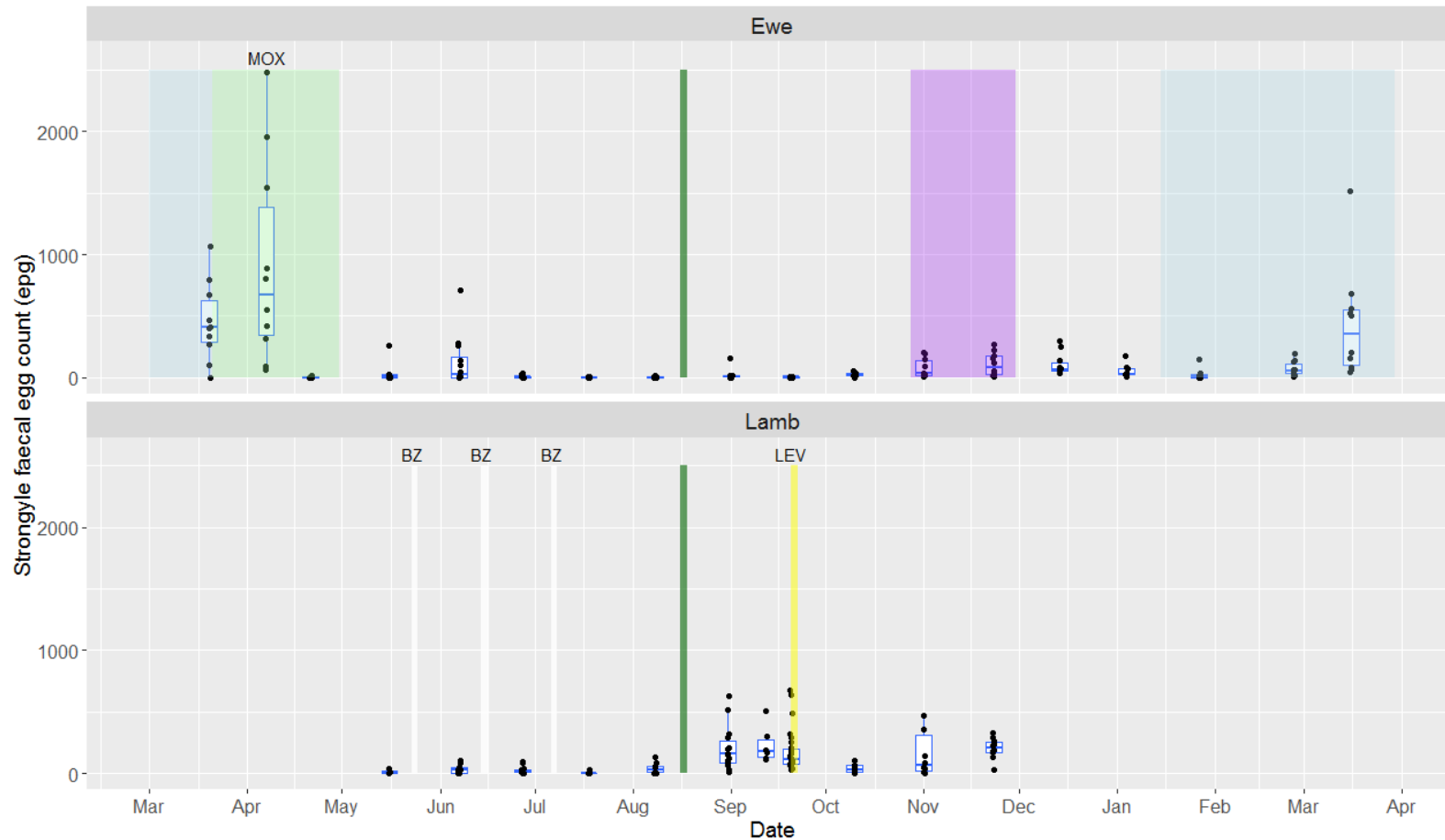
Lamb faecal samples were collected from 16<sup>th</sup> May 2016, from just four lambs in the initial sampling group, when lambs would have been eight weeks old or younger. The initial mean strongyle FEC was only 11.8 epg, but the presence of *N. battus* eggs were noted in the faeces (Figure 3.4, Table 3.4). Following this, ten lambs were sampled at each time point. Mean FECs remained low, before rising from 40.3 epg on the 8<sup>th</sup> August to 205.9 epg on the 31<sup>st</sup> August 2016 (median 30 epg rising to 157.5 epg, Table 3.4). Between these two dates the

lambs had been weaned, combined into one large group of approximately 300 lambs and moved to silage aftermath. Following this time point, lamb FECs remained higher, averaging between 151.8 and 232.5 epg until the end of November 2016, when very few lambs were left on the farm (Figure 3.4). Only one sample was low during this time period, with a mean 38.9 epg, collected on the 10<sup>th</sup> October 2016, 20 days after a LEV treatment was given (Table 3.4).

**Table 3.3: Farm 1 ewe strongyle faecal egg counts.**

Date	Number of ewes sampled	Arithmetic mean strongyle epg	SD strongyle epg	Range strongyle epg	Median strongyle epg
20/03/2016	10	452.1	318.9	(2-1062)	409.5
07/04/2016	10	912.0	822.7	(60-2484)	675.0
21/04/2016	10	3.4	5.7	(0-15)	0.0
16/05/2016	9	36.4	84.9	(0-261)	3.0
07/06/2016	12	129.8	208.7	(0-711)	30.0
27/06/2016	11	9.8	11.2	(0-36)	6.0
18/07/2016	10	2.8	4.0	(0-11)	1.0
08/08/2016	10	3.7	5.3	(0-16)	1.0
01/09/2016	13	20.3	43.2	(1-162)	5.0
20/09/2016	10	3.0	3.0	(0-8)	3.0
10/10/2016	10	23.0	16.0	(1-51)	19.5
01/11/2016	10	77.0	79.9	(4-207)	32.0
23/11/2016	10	106.2	95.3	(6-270)	84.0
14/12/2016	10	109.5	91.8	(39-297)	67.5
04/01/2017	9	53.0	53.2	(12-180)	30.0
27/01/2017	10	25.3	46.6	(1-153)	8.0
26/02/2017	10	75.9	59.1	(9-195)	58.5
16/03/2017	10	432.9	446.9	(42-1512)	354.0

Samples collected on the 7<sup>th</sup> and 21<sup>st</sup> of April were collected *per rectum* as part of a MOX FECRT. Epg = eggs per gram.



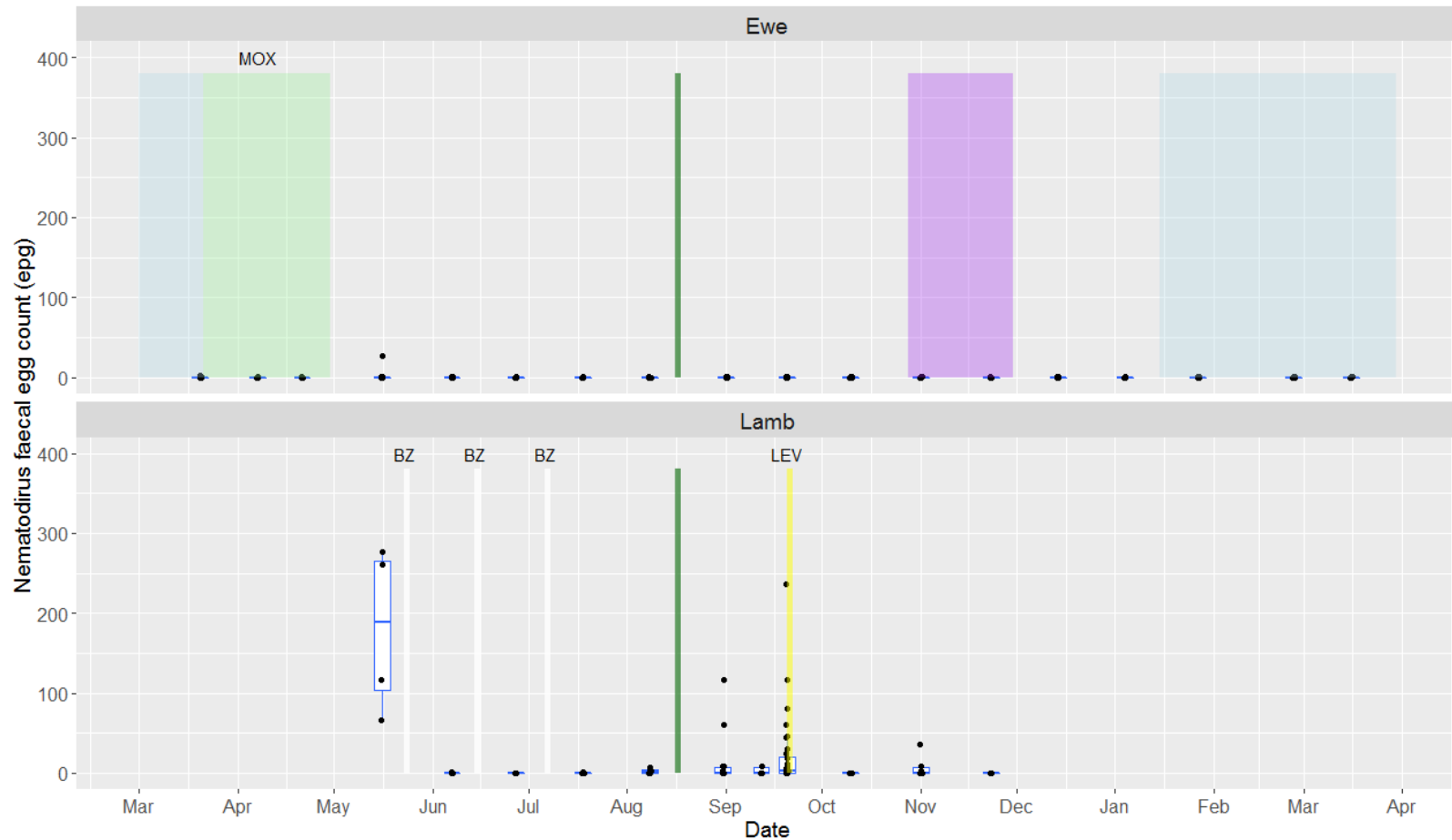
**Figure 3.4: Strongyle faecal egg counts over time on Farm 1. A box and whisker plot is shown for each sampling time point. In addition, each individual's egg per gram is plotted at each time point (dots). Key to annotations: light blue = ewes housed, light green = lambing, purple = breeding, dark green = weaning, white = albendazole (BZ) drench, yellow = LEV drench. Note, MOX indicates day zero of the MOX faecal egg count reduction test performed by the farm veterinary surgeons. At time of LEV treatment of remaining lambs, 32 lamb FECRTs were performed (post-treatment samples shown elsewhere).**

**Table 3.4: Farm 1 lamb strongyle faecal egg counts.**

Date	Number of lambs sampled	Arithmetic mean strongyle epg	SD strongyle epg	Range strongyle epg	Median strongyle epg
16/05/2016	4	11.8	15.1	(0-33)	7.0
07/06/2016	11	31.5	34.0	(0-99)	24.0
27/06/2016	14	25.1	30.4	(0-99)	13.0
18/07/2016	10	5.0	8.6	(0-24)	1.0
08/08/2016	10	40.3	42.6	(0-129)	30.0
31/08/2016	14	205.9	179.4	(9-630)	157.5
12/09/2016	6	232.5	149.2	(114-504)	175.5
20/09/2016	32	172.4	158.2	(27-672)	115.5
10/10/2016	10	38.9	32.3	(3-102)	28.5
01/11/2016	10	151.8	175.9	(2-468)	64.5
23/11/2016	10	203.7	86.9	(27-330)	205.5

Samples collected on the 20<sup>th</sup> September were collected *per rectum* as part of a faecal egg count reduction test, for which post-treatment results are shown separately. Epg = eggs per gram

*N. battus* eggs were noticed in ewe samples, but counts were always low to negligible (Figure 3.5). Of note was the presence of *N. battus* eggs in housed ewe samples on the 20<sup>th</sup> March 2016. *N. battus* eggs were found in the earliest lamb samples prompting treatment of the flock with albendazole (Figure 3.5, Table 3.5). This parasite continued to be detected at low levels in lamb faecal samples throughout the year, with an additional rise observed in September.



**Figure 3.5: *Nematodirus* faecal egg counts over time on Farm 1. A box and whisker plot is shown for each sampling time point. In addition, each individual's eggs per gram is plotted at each time point (dots). Key to annotations: light blue = ewes housed, light green = lambing, purple = breeding, dark green = weaning, white = albendazole (BZ) drench, yellow = LEV drench. Note, MOX indicates day zero of MOX faecal egg count reduction test performed by farm veterinary surgeons, at time of sampling. At time of LEV treatment of remaining lambs, 32 FECRTs were performed (post-treatment samples shown elsewhere).**

**Table 3.5: Farm 1 lamb *Nematodirus* faecal egg count.**

Date	Number of lambs sampled	Arithmetic mean <i>Nematodirus</i> epg	SD of <i>Nematodirus</i> epg	Range <i>Nematodirus</i> epg	Median <i>Nematodirus</i> epg
16/05/2016	4	180.0	104.5	(66-276)	189.0
07/06/2016	11	0.1	0.3	(0-1)	0.0
27/06/2016	14	0.0	0.0	(0-0)	0.0
18/07/2016	10	0.2	0.4	(0-1)	0.0
08/08/2016	10	2.3	2.6	(0-7)	1.5
31/08/2016	14	14.1	33.6	(0-117)	0.0
12/09/2016	6	3.0	4.6	(0-9)	0.0
20/09/2016	32	22.3	47.6	(0-237)	3.0
10/10/2016	10	0.0	0.0	(0-0)	0.0
01/11/2016	10	5.7	11.3	(0-36)	0.0
23/11/2016	10	0.0	0.0	(0-0)	0.0

Samples collected on the 20<sup>th</sup> September were collected *per rectum* as part of a faecal egg count reduction test, for which post-treatment results are shown separately. Epg = eggs per gram.

#### 3.3.1.4 Strongyle species present on the farm

Faecal examination identified 'strongyles', *Nematodirus*, *Strongyloides* (confirmed as a true infection by rectal sampling), *Trichuris*, *Monezia* and various species of *Coccidia*. Species identification of L3 strongyles by PCR revealed considerable variation and diversity within the strongyle population over time and between sheep age groups (Figure 3.6). The following strongyle species were identified: *Cooperia curticei*, *Chabertia ovina*, *Oesophagostomum venulosum*, *N. battus*, *T. circumcincta*, *Trichostrongylus vitrinus*, and *Trichostrongylus axei* (Figure 3.6). The proportions of these nematode species varied throughout the year and were generally different between the ewes and the lambs. It is important to note here that species proportions do not necessarily reflect abundance - species proportions when egg counts are high will have a very different impact to when egg counts are low.

Generally, either *T. circumcincta* or *C. curticei* predominated. Species common in the ewe samples were later seen to predominate in the lamb samples. However, usually for a given sampling time point, the majority species was different for ewes and lambs. Indeed, *C. curticei* was not apparent in the lamb samples until mid-September, despite being detected in pre-lambing samples



and accounting for over 63% of strongyles speciated in ewes in July (Figure 3.6). Taking species diversity into account, the epg attributable to individual pathogenic species, such as *T. circumcincta*, was often considerably lower than that calculated by the overall strongyle FEC (Figure 3.6).

In the ewes, *T. circumcincta* accounted for only 38 to 53% of strongyles speciated pre-lambing in 2016 (Figure 3.6). Post-MOX treatment, only four strongyle larvae were harvested post-coproculture, of which three were *T. circumcincta* and the other *N. battus*. In early June, coinciding with a small but noticeable rise in FEC, *T. circumcincta* accounted for 58% of strongyle L3 speciated, but in early July, when the FEC dropped to extremely low levels, *T. circumcincta* contributed only 4.6% in the ewe sample (Figure 3.6). Subsequent ewe FECs were low, with negligible proportions of *T. circumcincta*. Despite the rise in FEC in the autumn, the proportion of *T. circumcincta* did not increase, remaining low, and was not detected in a ewe sample collected in early January 2017 (Figure 3.6).

*Trichostrongylus* species were detected throughout the year, but were not particularly prevalent in any coproculture. Although *T. vitrinus* was always detected in the lamb samples, and most of the ewe samples, it was always a minor species (Figure 3.6). In the lambs, *T. axei* numbers were negligible and not detected until July. Nevertheless, *T. axei* had been present in the ewes pre-lambing and remained at detectable levels in all ewe samples across the year, increasing in proportion over the summer months, before declining in the autumn. In the ewe sample collected in early January 2017, *T. axei* accounted for 28.6% of the strongyles speciated by PCR (Figure 3.6).

Two large intestinal species were apparent over time in both the ewes and the lambs. *C. ovina* was present in the ewe samples at 29.0% and 31.4% pre-lambing and in the July lamb sample at 37.7%. Only one ewe sample did not include *C. ovina*. Lastly, *O. venulosum* was present in ewe samples pre-lambing, although interestingly did not re-appear in either ewe or lamb samples until late August/early September. However, it was a minor species, so may have been present at low levels, undetectable by the methods used in this study.

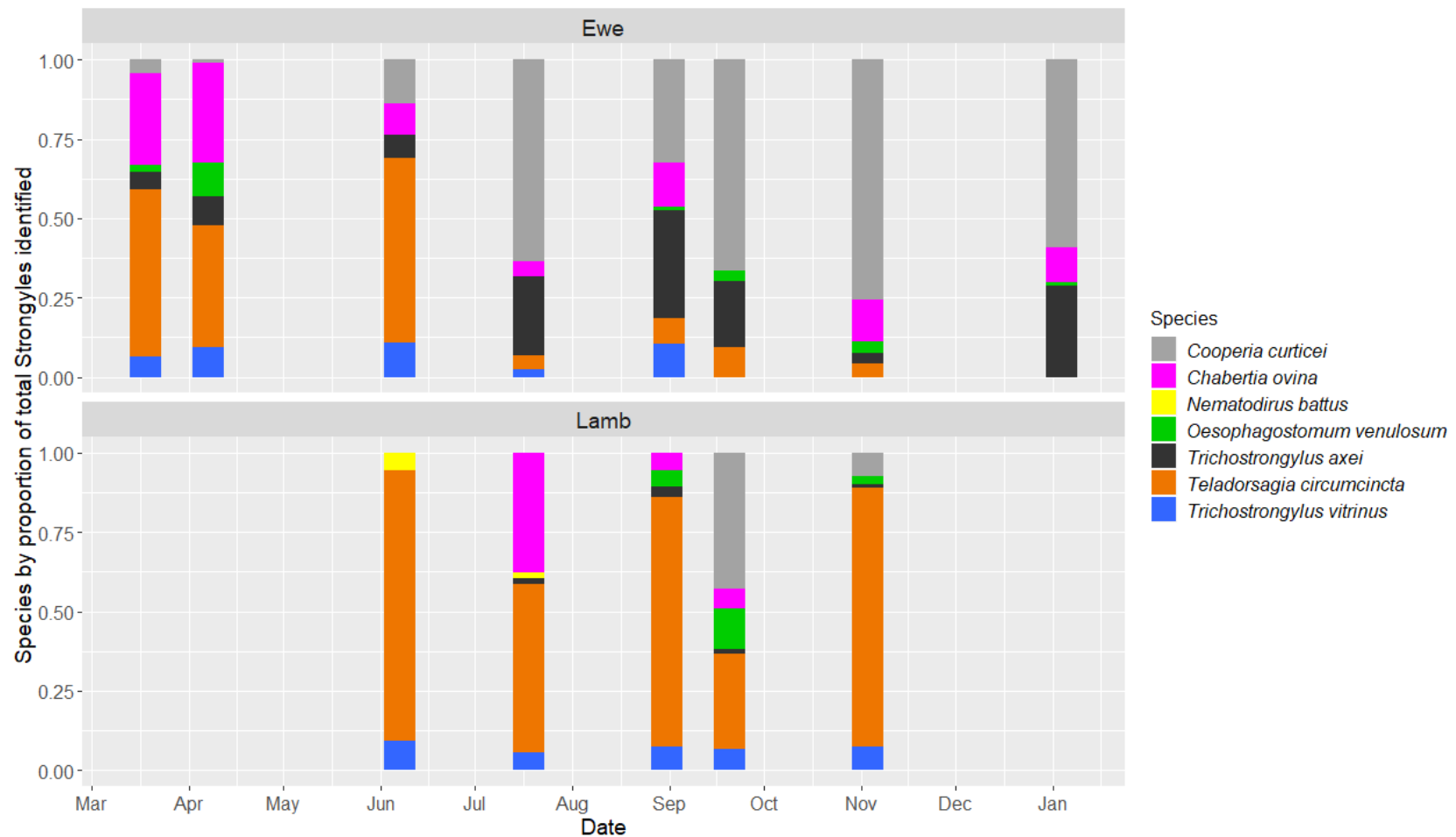


Figure 3.6: Species proportions of cultured strongyle larvae on Farm 1 over time. From left to right the number of total strongyles speciated at each time point was; Ewes: 93, 86, 93, 88, 86, 63, 91 and 91; Lambs: 54, 53, 93, 222, 81. Note that 20<sup>th</sup> September 2016 is an average of samples cultured and speciated separately from the three lamb pre-treatment samples from the faecal egg count reduction test. Those speciated post-treatment are shown later (Figure 3.7).

### 3.3.2 Determining anthelmintic efficacy using available tests

#### 3.3.2.1 Faecal egg count reduction test

In September 2016, a FECRT was performed drawing from a group of lambs initially chosen at random by the farmer. Based on those lambs which were caught first, but with the aim of roughly balancing for sex, ten, ten and twelve lambs were chosen. The anthelmintic classes tested were 1-BZ, 2-LV and 3-ML respectively (see 2.2.5). FECs were performed on days zero, seven and fourteen post-treatment. Resistance to all three anthelmintics (albendazole, LEV and IVM) was detected (Table 3.2). FECs suggested an overall efficacy of 65% (95% HPD Interval: 14.5, 86.2) for albendazole, 80.8% efficacy (95% HPD Interval: 58.0, 91.1) for LEV and 77% efficacy (95% HPD Interval: 45.5, 91.6) for IVM (Table 3.2).

Pre-treatment, strongyle speciation by PCR was performed and a diverse mix of strongyle species was detected (Figure 3.6 and Figure 3.7). Post-treatment, molecular speciation was performed for albendazole and IVM at days seven and fourteen. Only *T. circumcincta* was detected post-IVM, however *T. circumcincta* and *C. curticei* were detected on day seven post-BZ, with *T. axei* and *O. venulosum* also identified on day 14 (Figure 3.7). Post-LEV, from the day seven sample, larvae were selectively picked under low power to obtain *T. circumcincta* individuals for NGS work, and too few larvae were recovered in total to also permit an un-biased speciation of the strongyle population. Adjustment of the FEC to represent the proportion of larvae identified as *T. circumcincta* indicated no change in the egg output of this species following treatment by either BZ or IVM (Table 3.6, Figure 3.8). *C. curticei* epg had fallen to 5 epg post-BZ from an initial 148 epg on day 0 (Table 3.6, Figure 3.8). Overall, there was a significant reduction in egg output from strongyle species other than *T. circumcincta* ( $p = 0.00$ ).

Table 3.6: Benzimidazole and ivermectin faecal egg count reduction tests performed in September 2016 on Farm 1.

Sample ID	N <sup>1</sup>	Mean Strongyle epg	SD Strongyle epg	Mann-Whitney test Strongyle epg <sup>2</sup>		Proportional <i>Teladorsagia circumcincta</i> Mean epg <sup>3</sup>	SD <i>Teladorsagia circumcincta</i> epg <sup>3</sup>	Mann-Whitney test <i>Teladorsagia circumcincta</i> epg <sup>3</sup>		Strongyle FECR % (95% HPD Interval) <sup>4</sup>	<i>Teladorsagia circumcincta</i> FECR % (95% HPD Interval) <sup>4</sup>	<i>Cooperia curticei</i> FECR % (95% HPD Interval) <sup>4</sup>
				D7	D14			D7	D14			
BZ D0	10	209	182	<b>p=0.005</b>	<b>p=0.005</b>	41.3	36	p=0.056	p=0.212			
BZ D7	8	47.1	75.4		p=0.534	21.3	34		p=0.625			
BZ D14	10	77	156			67	137			65.3% (14.5, 86.2%)	4.99% (0.0, 55.8%)	91.8% (82.4, 96.1%)
IVM D0	12	202	186	<b>p=0.003</b>	<b>p=0.002</b>	51	47	p=0.976	p=0.583			
IVM D7	11	53.8	53.6		p=0.666	53.8	53.6		p=0.666			
IVM D14	12	48.4	57.9			48.4	57.9			77.2% (45.5, 91.6%)	9.17% (0.0, 53.2%)	

Arithmetic mean strongyle egg counts and proportional *Teladorsagia circumcincta* egg counts (adjusted by species percentage as determined by PCR) are reported, with strongyle faecal egg count reduction (FECR) percentages for each anthelmintic. The BZ FECR percentage has been calculated for both *T. circumcincta* and *Cooperia curticei* but only for *T. circumcincta* post-IVM treatment as all other species egg counts were zero on day 14 post-IVM. <sup>1</sup>N = number of lambs sampled. On day 7 not all lambs were present or contained sufficient material *per rectum* for an adequate sample to be obtained. <sup>2</sup>Comparing egg count medians between sampling dates using Minitab; hypothesis of 'medians not equal', p value reported. Numbers in bold indicate significant differences between sample medians. <sup>3</sup>*T. circumcincta* epg calculated for each individual sample using the strongyle species proportions obtained from PCR of the ITS2 region of L3 (D0, D7) and eggs/L1 (D14). <sup>4</sup>FECR percentage calculated using R shiny 'eggCounts' (Torgerson et al., 2014; Wang and Paul, 2017; Wang et al., 2017). Bayesian CI (HPD.Low and HPD.High) are in brackets. HPD = Highest posterior density interval. Epg = eggs per gram.

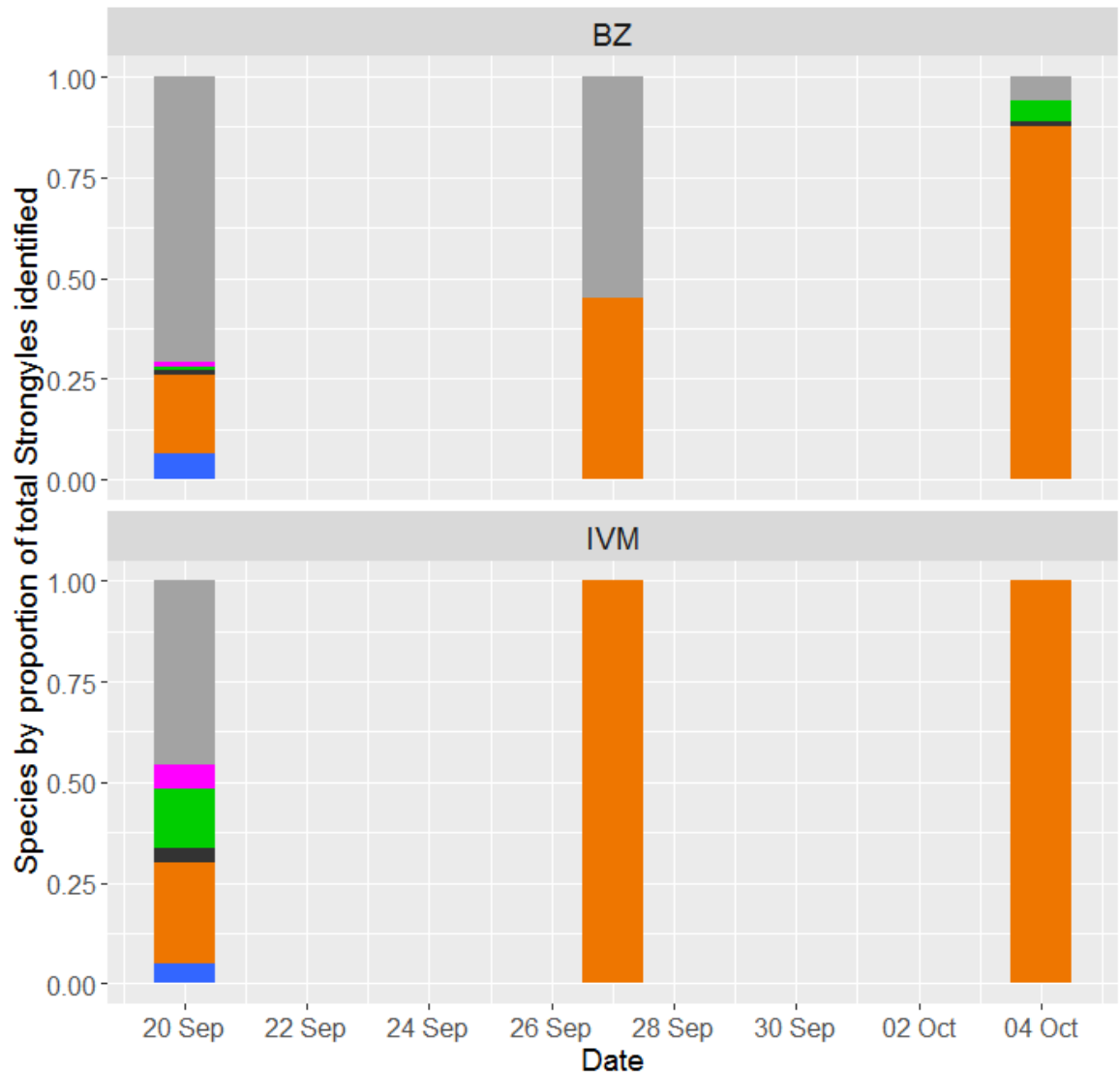
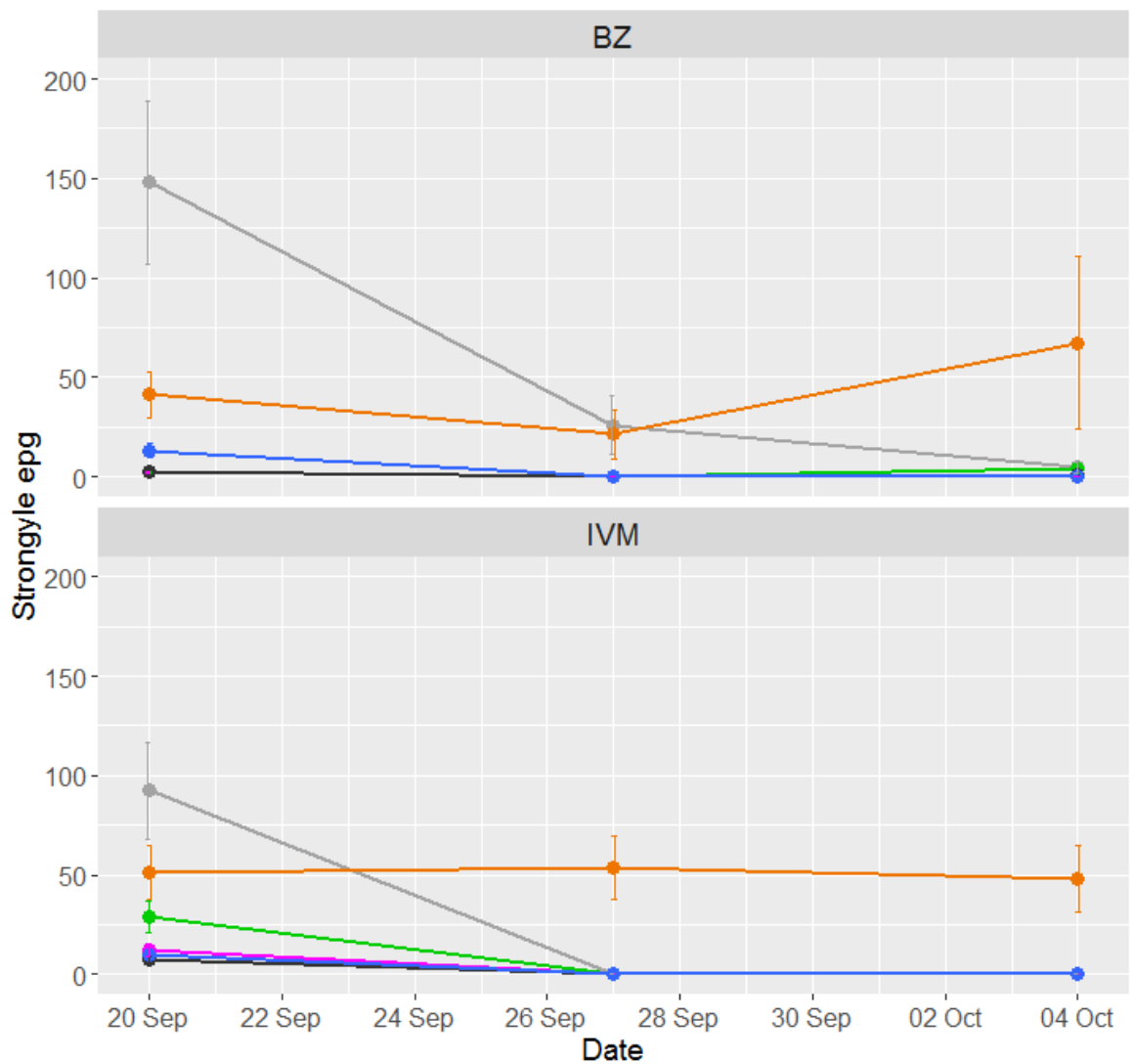


Figure 3.7: Farm 1 benzimidazole and ivermectin faecal egg count reduction test strongyle species proportions. The proportions of strongyles as speciated by PCR of the ITS2 region are shown for each sampling time point (days 0, 7 and 14 of each FECRT). Key to colours: grey = *Cooperia curticei*, pink = *Chabertia ovina*, green = *Oesophagostomum venulosum*, black = *Trichostrongylus axei*, orange = *Teladorsagia circumcincta*, blue = *Trichostrongylus vitrinus*. From left to right the following number of total strongyles was speciated at each time point; BZ: 96, 62, 82; IVM: 83, 83, 42. L3 were used on days 0 and 7 but on day 14, individuals sampled were eggs and L1 larvae from the wells of an egg hatch test (in equal proportions to the percentage hatch).



**Figure 3.8: Farm 1 benzimidazole and ivermectin faecal egg count reduction test mean strongyle eggs per gram by species proportion. Individual lamb faecal egg counts were extrapolated by L3 strongyle species proportions identified. The arithmetic mean for each species is shown, with the standard error of the mean plotted. Key to colours: grey = *Cooperia curticei*, pink = *Chabertia ovina*, green = *Oesophagostomum venulosum*, black = *Trichostrongylus axei*, orange = *Teladorsagia circumcincta*, blue = *Trichostrongylus vitrinus*. From left to right the following number of total strongyles was speciated at each time point; BZ: 96, 62, 82; IVM: 83, 83, 42. L3 were used on days 0 and 7 but on day 14, individuals sampled were eggs and L1 larvae from the wells of an egg hatch test (in equal proportions to the percentage hatch).**

### 3.3.2.2 Bioassays

Anthelmintic resistance can also be detected using *in vitro* bioassays, testing sensitivity to the anthelmintic by free-living lifecycle stages (Coles et al., 2006). Throughout the season, samples were tested using an EHT for BZ resistance and a LDT for IVM resistance. The former used thiabendazole (TBZ) and the latter 22,23 dihydroavermectin B1 (IVM<sub>B1a</sub>). Both were liquid tests, the former following the protocol defined by von Samson-Himmelstjerna et al. (2009a) written after

ring-testing in European labs, and the latter was adapted from work by Varady et al. (1996) (see 2.3.2).

These bioassays were used both pre- and post- the BZ and IVM FECRTs performed in September 2016 on the farm to correlate the *in vivo* phenotype observed with the *in vitro* phenotype of the offspring. For the EHT, the percentage hatch in the DMSO control wells showed a difference between the pre- and post-BZ treatment strongyle populations. Pre-BZ, 79% of eggs hatched in control wells compared to 91% post-BZ (Table 3.7). After adjustment for the response in the DMSO control wells, it was determined that just 39% of eggs from the pre-BZ strongyle sample hatched in the EHT at the definitive dose of 0.1 µg/ml (Coles et al., 2006). Post-BZ, 93% hatched (Table 3.8), giving a resistance ratio (RR) of 2.4 (% hatch post-treatment/% hatch pre-treatment). Calculating the ED<sub>50</sub> also revealed a difference pre- and post-BZ treatment. Post-BZ, 50% hatched at a higher concentration of TBZ/DMSO (0.615 µg/ml) compared with pre-BZ (0.048 µg/ml) (Table 3.8, Figure 3.9), a RR of 12.8 (ED<sub>50</sub> post-treatment/ED<sub>50</sub> pre-treatment). The pre-IVM strongyle sample was similar to the pre-BZ sample, with an ED<sub>50</sub> of just 0.043 µg/ml (Table 3.8, Figure 3.10), and a percentage hatch in the 0.1 µg/ml definitive dose well of 32.6% (Table 3.8). Post-IVM, not enough material was available to perform both a full EHT and a full LDT, and so only the definitive dose well was assessed post-IVM, with no control well used. Though imperfect, this still revealed an uncorrected percentage hatch of 82.4% post-IVM (Table 3.8).

**Table 3.7: Control wells of the egg hatch tests performed with samples taken during the faecal egg count reduction test .**

Test	Average percentage hatch
	0.5% DMSO
Pre-BZ FECRT (4 plates)	79%
Post-BZ FECRT	91%
Pre-IVM FECRT (4 plates)	85%

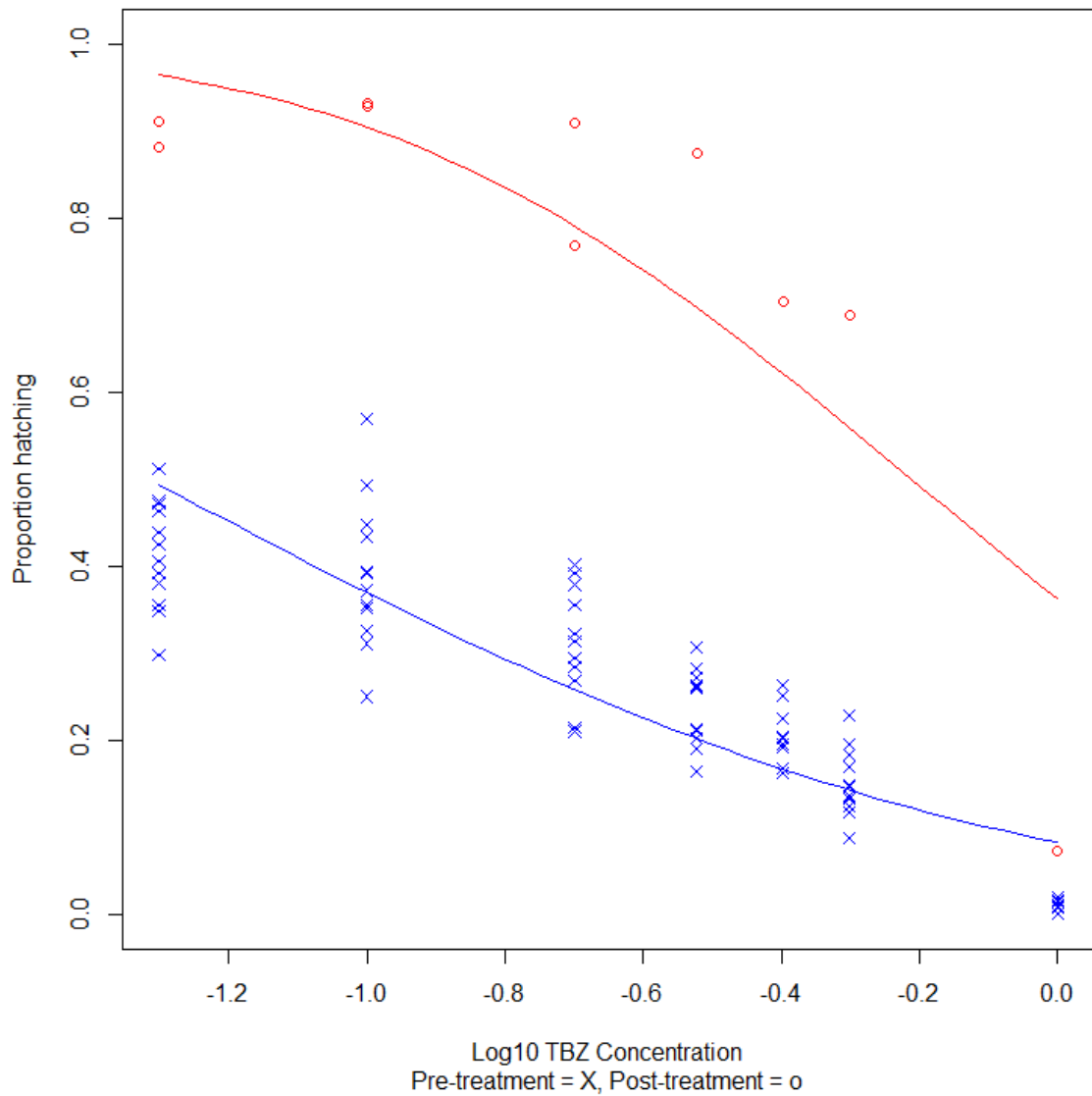
The average percentage hatch in the 0.5% DMSO control wells is reported for each sample population tested.

**Table 3.8: Egg hatch test results.**

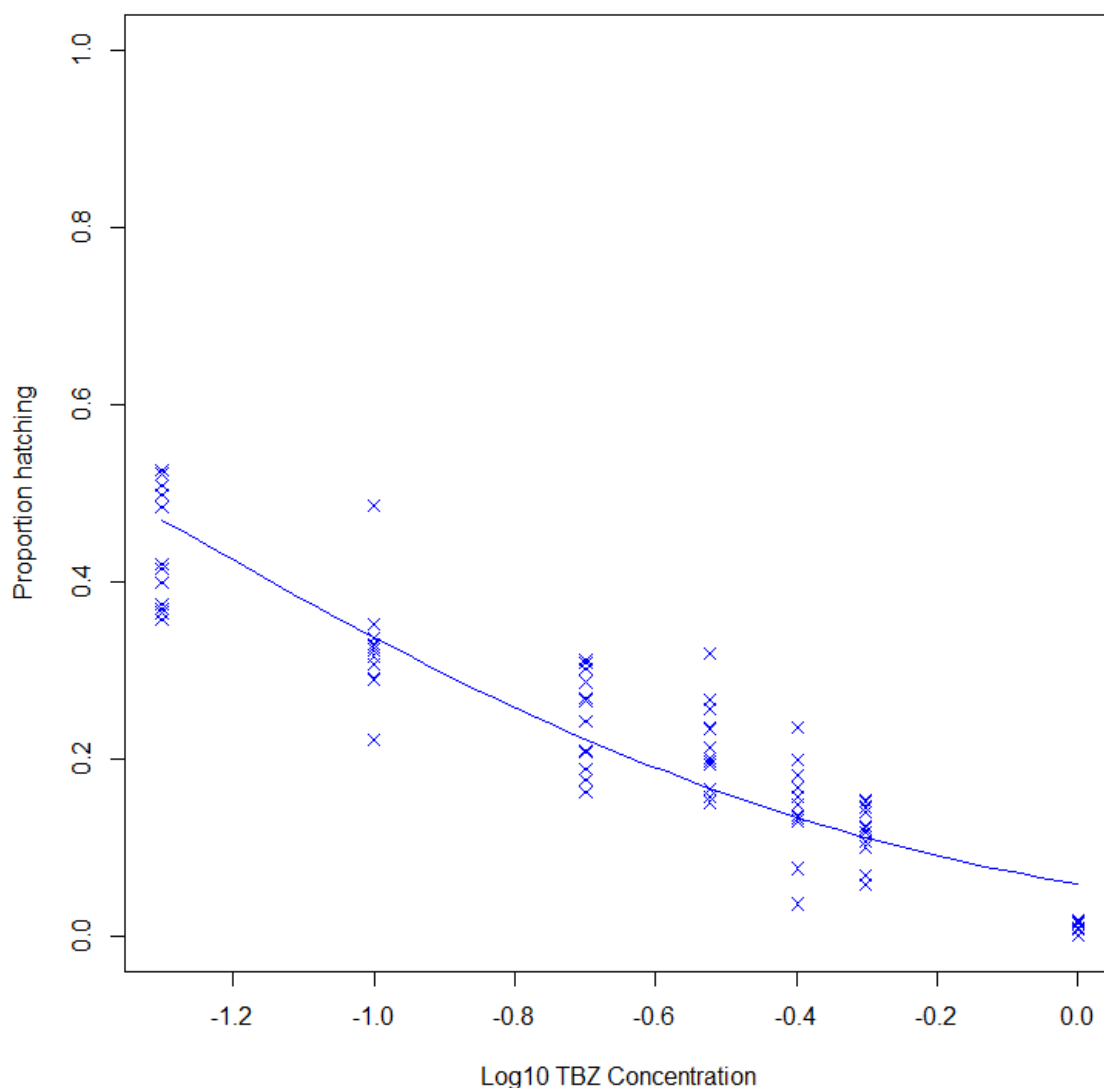
Test	EHT ED <sub>50</sub> (µg/ml) <sup>1</sup> (95% CI)	Percentage hatch in 0.1 µg/ml <sup>2</sup> (SD)
Pre-BZ FECRT (four plates combined analysis)	0.048 µg/ml (0.045, 0.052)	39.1% (8.58)
Post-BZ FECRT	0.615 µg/ml (0.538, 0.730)	93.1% (0.23)
Pre-IVM FECRT (four plates combined analysis)	0.043 µg/ml (0.040, 0.046)	32.6% (6.06)
Post-IVM FECRT	NA	82.4% <sup>3</sup> (3.43)

Effective doses (ED<sub>50</sub>) of thiabendazole required to prevent hatching of 50% of eggs are reported for each full test performed. The average percentage hatch in definitive dose wells (0.1 µg/ml) is given for each test. <sup>1</sup>Raw data were adjusted for response in 0.5% DMSO control wells. <sup>2</sup>Adjusted for response in 0.5% DMSO control wells. <sup>3</sup>No DMSO well was included post-ivermectin and only two replicate wells for 0.1 µg/ml were included. This therefore is the uncorrected percentage hatch. The uncorrected percentage hatch in the 0.1 µg/ml wells pre-ivermectin was 27.52% (5.0 SD).





**Figure 3.9: Dose response curve of the egg hatch test performed using samples from the benzimidazole faecal egg count reduction test. The corrected egg hatch test data are shown, with  $\log_{10}$ (thiabendazole concentration) plotted against the proportion hatching. The data were modelled in R using a binomial general linear model (probit), and the regression line is plotted here. Pre-BZ (X, blue), four tests were performed, each with three replicate wells for each concentration included. Each well is plotted. On day 14 post-BZ, lambs were sampled and a further EHT performed. Fewer eggs were available and for some concentrations only one or two replicate wells were included. The proportion hatching in each well is shown (O, red).**



**Figure 3.10: Dose response curve of the egg hatch test performed using the Pre-ivermectin faecal egg count reduction test sample. The corrected egg hatch test data are shown, with  $\log_{10}$ (thiabendazole concentration) plotted against the proportion hatching. The data were modelled in R using a binomial general linear model (probit), and the regression line is plotted here. Pre-treatment, four tests were performed, each with three replicate wells for each concentration included. Each well is plotted. Post-ivermectin data are not shown.**

Average egg hatch in the 2% DMSO control wells of the LDT was 97%. Pre-treatment 67% of hatched larvae developed to L3 on average in the control wells (range 53-75%), compared to 86% post-IVM (Table 3.9). No definitive dose has been established for the IVM LDT used in this study, but using a GLM, the  $ED_{50}$  can be determined. Pre-IVM this was 2.43 nM (95% CI: 2.14, 2.81) and post-IVM this rose to 4.07 nM (95% CI: 2.87, 6.36) (Table 3.10, Figure 3.11), giving a RR of 1.7. Pre-BZ treatment an  $ED_{50}$  of 2.27 nM was calculated, comparable to that of the pre-IVM group (Table 3.10, Figure 3.12).

**Table 3.9: Control wells of the larval development test.**

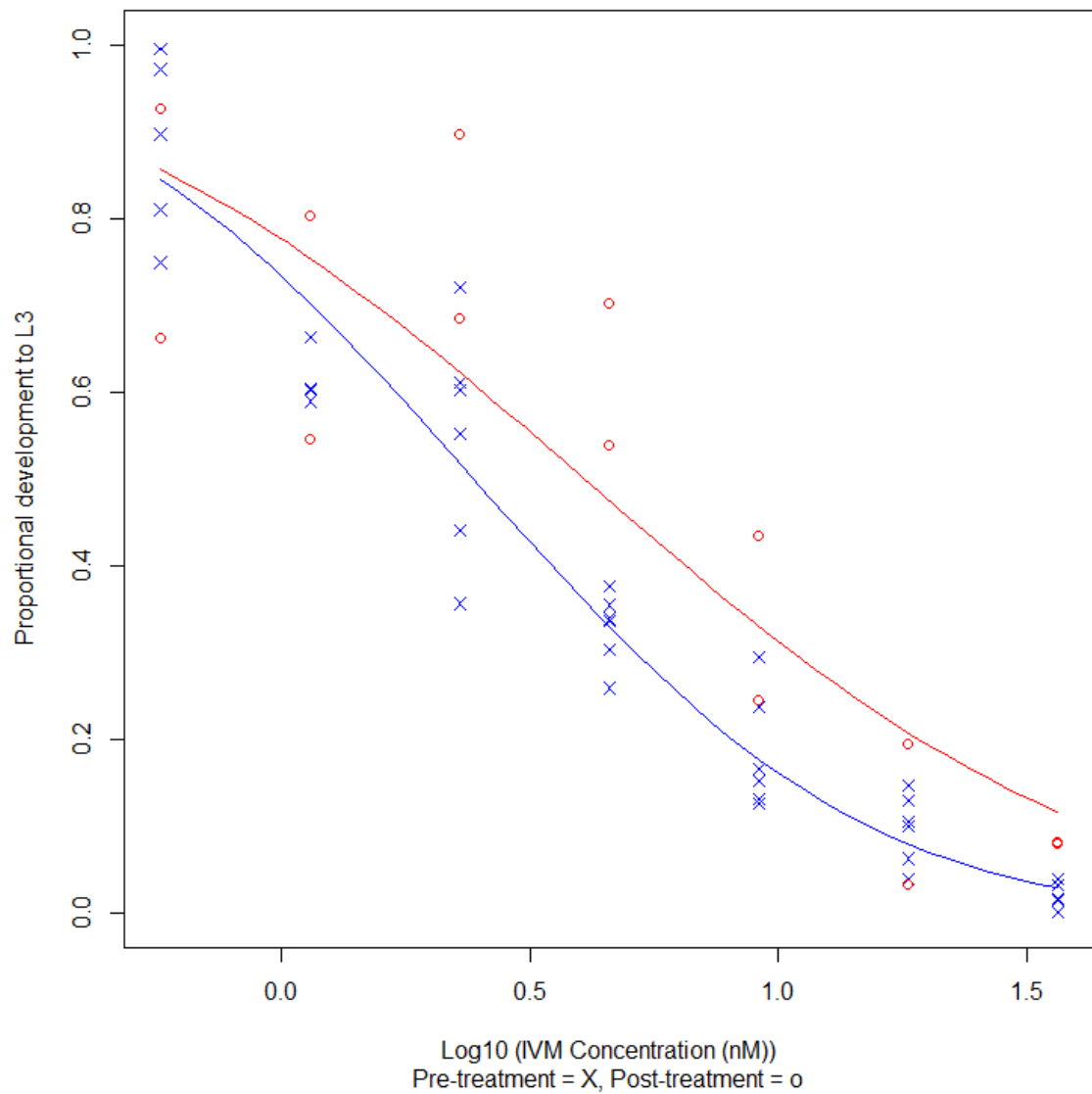
Test	Average percentage hatch		Average percentage development to L3	
	Water	2% DMSO	Water	2% DMSO
Pre-BZ FECRT, plate 1	98%	97%	87%	75%
Pre-BZ FECRT, plate 2	98%	97%	85%	67%
Pre-IVM FECRT, plate 1	99%	98%	78%	58%
Pre-IVM FECRT, plate 2	96%	96%	66%	53%
Post-IVM FECRT	95%	92%	92%	86%

The average percentage hatch and development to L3 are shown for comparison between the water and 2% DMSO control wells. Samples used to perform the test were obtained during faecal egg count reduction tests of a benzimidazole anthelmintic (BZ) and ivermectin (IVM).

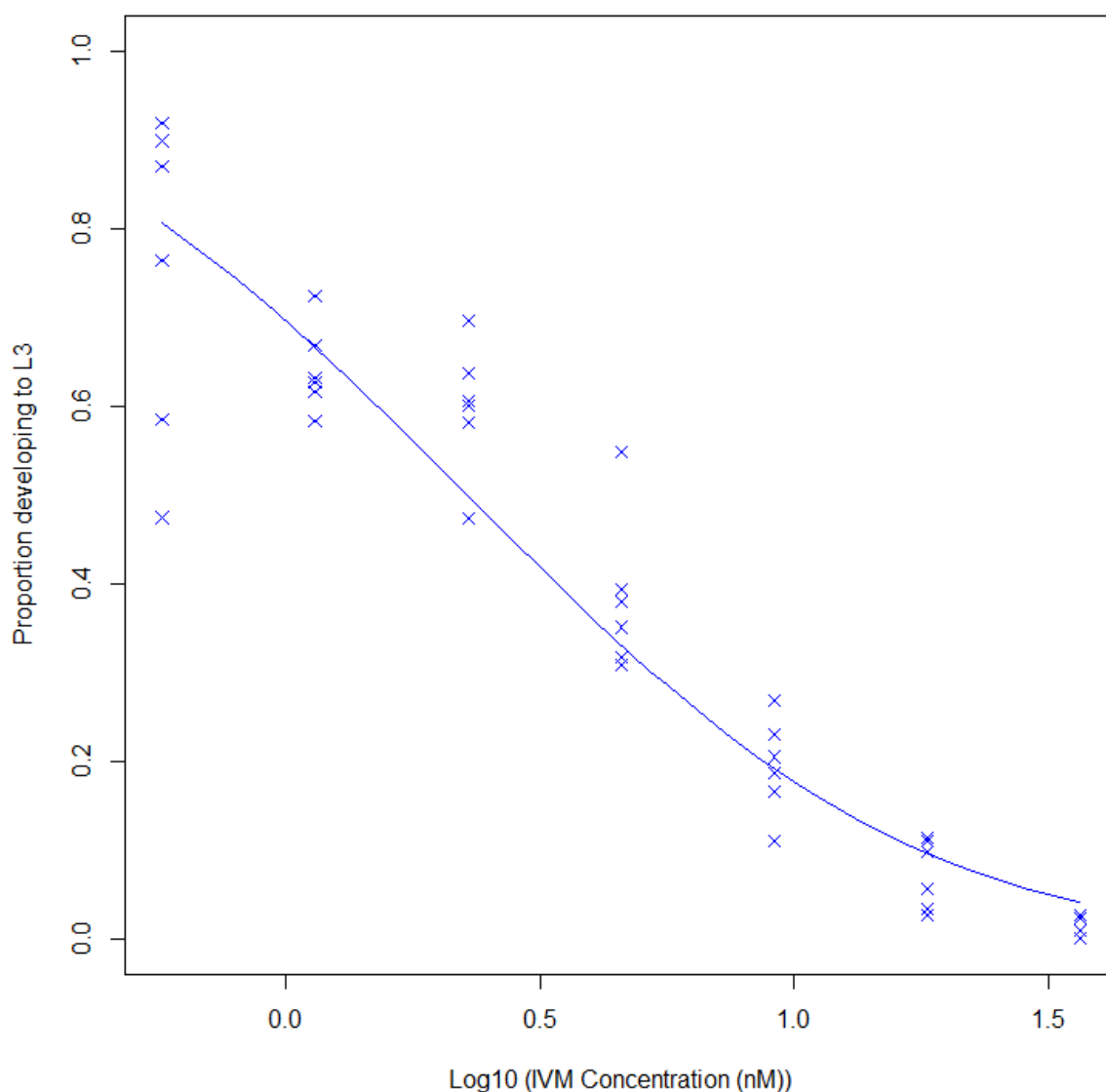
**Table 3.10: Larval development test results.**

Test	IVM LDT ED <sub>50</sub> (nM) <sup>1</sup> (95% CI)
Pre-BZ FECRT (two plates combined analysis)	2.27 nM (2.05, 2.56)
Pre-IVM FECRT (two plates combined analysis)	2.43 nM (2.14, 2.81)
Post-IVM FECRT	4.07 nM (2.87, 6.36)

Effective doses (ED<sub>50</sub>) of ivermectin required to inhibit 50% development of eggs to L3 are reported for each test performed. <sup>1</sup>Raw data were adjusted for response in 2% DMSO control wells. Samples used to perform the test were obtained during faecal egg count reduction tests of a benzimidazole anthelmintic (BZ) and ivermectin (IVM).



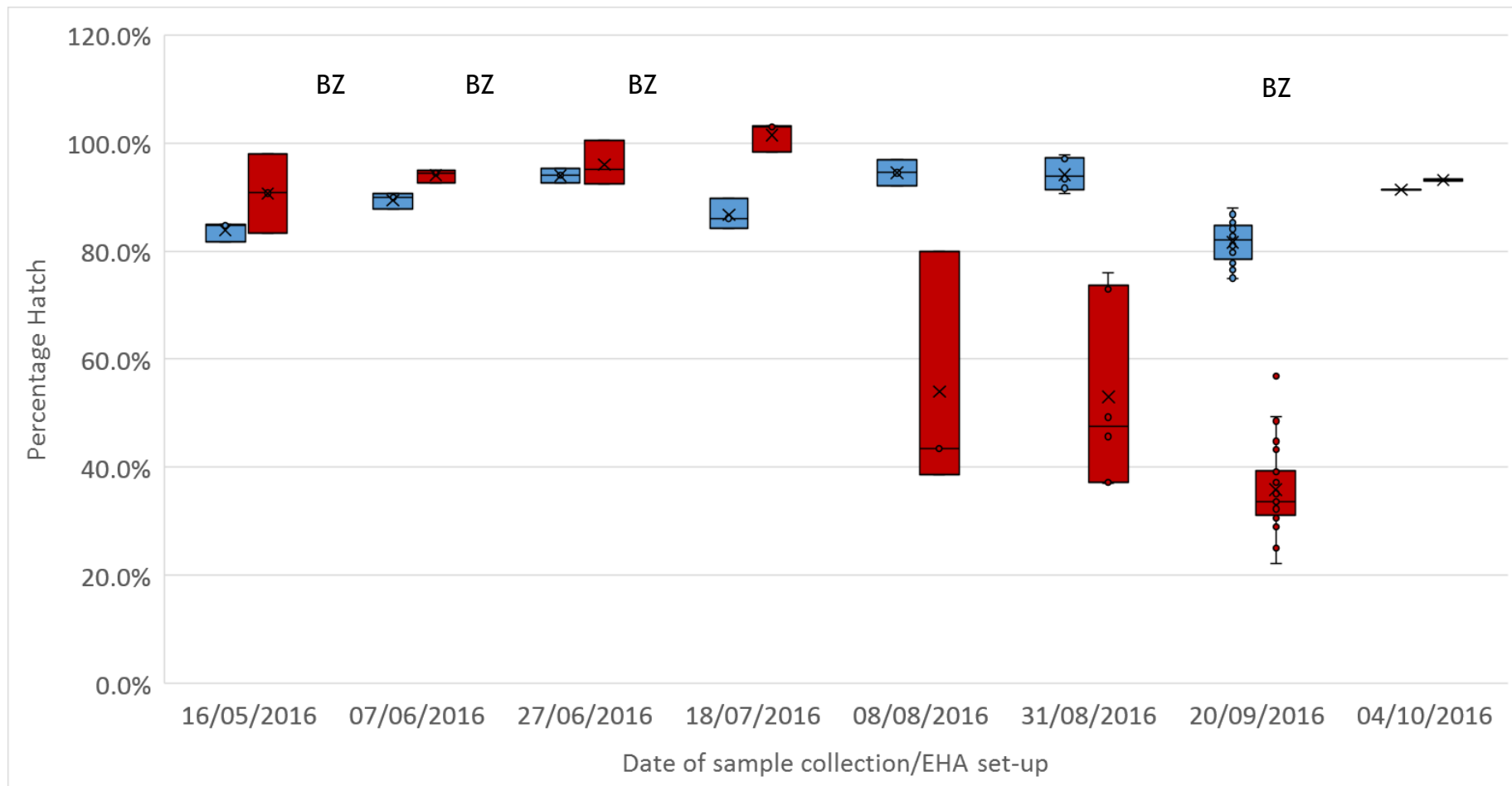
**Figure 3.11: Dose response curve of the larval development test performed using samples obtained during an ivermectin faecal egg count reduction test. The corrected larval development test data are shown, with  $\log_{10}$ (ivermectin concentration) plotted against the proportion developing to L3. The data were modelled in R using a binomial general linear model (probit), and the regression line is plotted here. Pre-treatment (X, blue), two tests were performed, each with three replicate wells per concentration included. Each well is plotted. On day 14 post-ivermectin (O, red), fewer eggs were harvested and concentrations were only performed in duplicate.**



**Figure 3.12: Dose response curve of the larval development test performed using the Pre-benzimidazole faecal egg count reduction test sample. The corrected larval development test data are shown, with  $\log_{10}$ (ivermectin concentration) plotted against the proportion developing to L3. The data were modelled in R using a binomial general linear model (probit), and the regression line is plotted here. Pre-treatment (X, blue), two tests were performed, each with three replicate wells for each concentration included. Each well is plotted. No larval development test was carried out post-benzimidazole treatment due to a lack of eggs.**

Over the season the EHT was used to assess the BZ sensitivity of the lamb strongyle population. For each sample tested, the hatch percentage in the DMSO control wells, and the corrected hatch percentage of the definitive dose wells (0.1  $\mu\text{g}/\text{ml}$ ) are shown in Figure 3.13. From May to July, percentage hatch in both the control wells and the definitive dose wells (0.1  $\mu\text{g}/\text{ml}$ ) averaged greater than 80%, and variation between replicate wells within a test was minimal. However, in August and September within-test variation increased considerably,

especially at the definitive dose. Lower percentage hatch was recorded in August and September at the definitive dose, compared with earlier in the year (Figure 3.13). In contrast, in October, 14 days following BZ treatment, both the egg hatch in the control well and the definitive dose wells were very similar, and hatching percentages were similar to those in early summer (Figure 3.13).



**Figure 3.13: Lamb egg hatch test results over the season. The plot shows the percentage hatch in the DMSO control wells (blue, uncorrected) and the definitive dose wells (0.1  $\mu\text{g/ml}$ , corrected by average percentage hatch in control wells, red) of egg hatch tests performed at each sampling time point. Number of tests performed (each with wells included in triplicate) at each sampling time point from 16<sup>th</sup> May 2016 onwards: 1, 1, 1, 1, 1, 2, 8, 1 (the latter had just one DMSO well and two definitive dose wells). Individual points are shown (each point represents one well), and the X indicates the mean value. BZ = Lambs treated with benzimidazoles.**

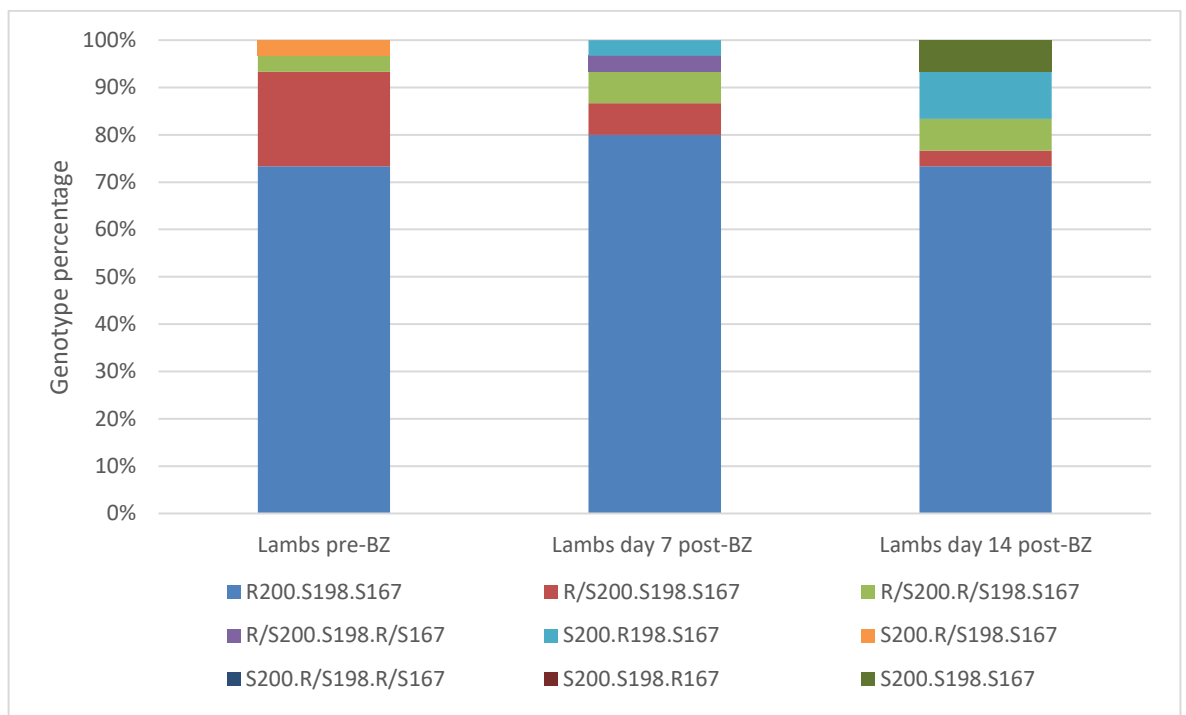
### 3.3.2.3 Pyrosequencing

The use of a molecular test, where a genetic marker for resistance is known, enables the rapid and accurate detection of resistance or sensitivity to an anthelmintic. BZ resistance is associated with three SNPs within the  $\beta$ -tubulin isotype-1 gene in many strongyle species (Kwa et al., 1995; Silvestre and Cabaret, 2002; Ghisi et al., 2007), and various genetic tests are available to detect these SNPs. For this study, a pyrosequencing method, developed by Skuce et al. (2010) and subsequently used by Redman et al. (2015) in field populations, was used to measure BZ resistance on this farm. Over the year, 386 individual *T. circumcincta* L3 were sequenced at the 198 and 200 SNPs (single sequencing reaction) and 385 L3 at the 167 SNP (one individual did not amplify). Those classified as ‘fully susceptible’ encoded only phenylalanine at codons 200 and 167 (P200F, P167F) and only glutamic acid at codon 198 (P198E). Individuals referred to here as ‘resistant’ to BZ carried at least one resistance associated SNP at either position 200, 198 or 167, encoding tyrosine at positions 200 or 167 (TTC>TAC), and leucine at codon 198 (GAA>TTA). Heterozygous individuals were identified with codon changes at more than one position. However, no individual was sequenced with three codon changes.

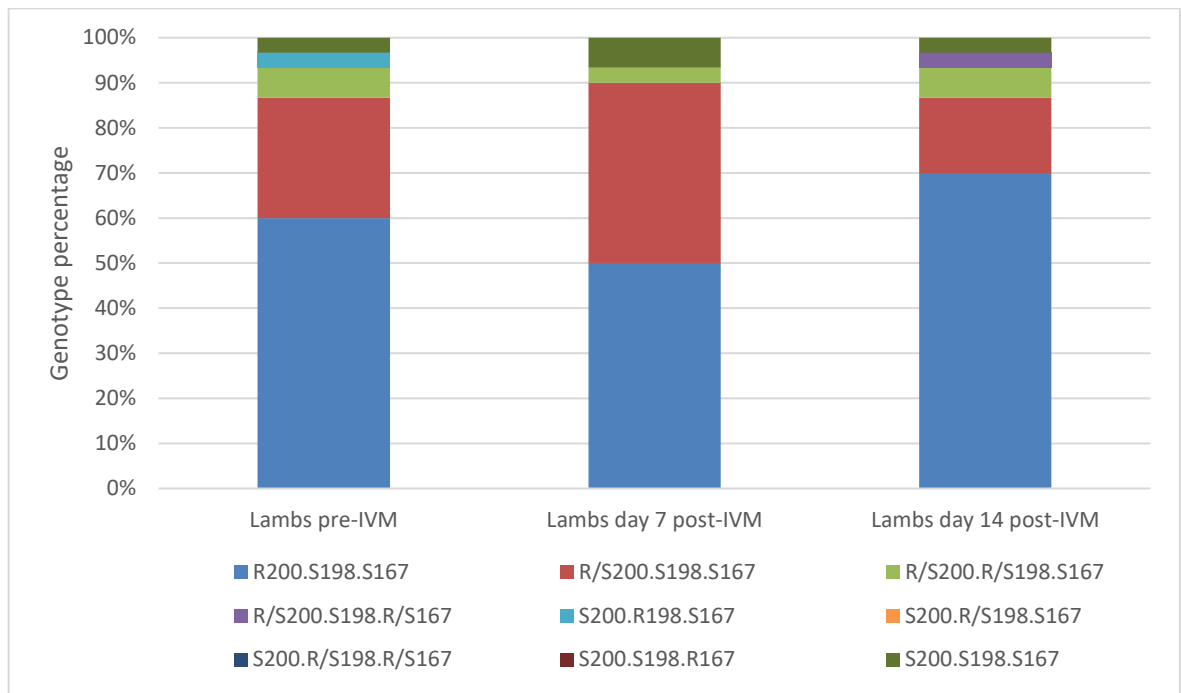
Of the 386 individuals sequenced, 64% were homozygous for the resistance associated SNP at codon 200, 28.5% were heterozygous at codon 200 and only 7.5% were homozygous for phenylalanine at this position. Of the 110 individuals heterozygous at codon 200, five (4.5%) were also heterozygous at codon 167, and 16 (14.5%) were heterozygous at codon 198. Of the 29 individuals homozygous susceptible at codon 200, eight (27.6%) were homozygous for the resistance associated SNPs at codon 198, four (13.8%) were heterozygous at codon 198 and one (3.4%) was heterozygous at codon 167. Only 16 (4.1%) individuals were fully susceptible at all three positions. In total 2.1% of all 386 individuals were homozygous for the resistance associated SNPs at codon 198, and 5.2% of all individuals were heterozygous at codon 198. All individuals with resistance-associated polymorphisms at codon 198 encoded leucine. In addition, one individual encoded a stop codon at position 198 (GAA > TAA). Only 6 individuals (1.6%) carried a resistance associated SNP at codon 167 and all were heterozygous.



Included in these 386 individuals sequenced were 30 *T. circumcincta* individuals selected at each sampling time point of the BZ and IVM FECRTs in September 2016, totalling 180 individuals (Figure 3.14 and Figure 3.15). Individuals homozygous for tyrosine at codon 200 or heterozygous at one or more codon were detected in all samples. Those with homozygous resistance mutations at codon 198 were detected on day 0 of the IVM FECRT and on days 7 and 14 of the BZ FECRT. ‘Susceptible’ individuals were identified on days 0, 7 and 14 of the IVM FECRT and on day 14 of the BZ FECRT. For both the BZ and IVM treatment groups no significant differences in genotype proportions were noted between samples before and after treatment (Figure 3.14 and Figure 3.15).



**Figure 3.14: *Teladorsagia circumcincta* offspring genotypes determined by pyrosequencing before and after benzimidazole treatment of lambs. At each sampling time point (day 0, 7 and 14 of a benzimidazole faecal egg count reduction test), 30 *T. circumcincta* individuals were genotyped at the 200, 198 and 167 codons of the  $\beta$ -tubulin isotype-1 gene. R = resistant allele SNP, S = sensitive allele SNP. R/S = heterozygous.**



**Figure 3.15: *Teladorsagia circumcincta* offspring genotypes determined by pyrosequencing before and after ivermectin treatment of lambs. At each sampling time point (day 0, 7 and 14 of an ivermectin faecal egg count reduction test), 30 *T. circumcincta* individuals were genotyped at the 200, 198 and 167 codons of the  $\beta$ -tubulin isotype-1 gene. R = resistant allele SNP, S = sensitive allele SNP. R/S = heterozygous.**

The total length of the sequence containing the 198 and 200 codons was 16 basepairs and contained additional, synonymous SNPs surrounding the non-synonymous SNPs associated with BZ resistance. Pyrosequencing allows determination of the individual's genotype at each basepair position but does not allow phasing of haplotypes for each individual. Taking a highly conservative approach - only fully homozygous individuals, or those with no more than one heterozygous SNP (i.e. excluding those heterozygous at both codon 200 and codon 198), and ignoring the codon 167 sequence, haplotypes were examined in 217 individuals, providing a total of 434 haplotypes (assuming two copies of the  $\beta$ -tubulin isotype-1 gene were amplified). Three 'susceptible' haplotypes, two 'resistant' haplotypes encoding leucine at codon 198 (P198L, P200F), and five 'resistant' haplotypes encoding tyrosine at codon 200 (P198E, P200Y) were identified. Three of the five haplotypes encoding tyrosine were commonly sequenced (88 to 158 sequence copies per haplotype), with two sequenced only five times each. One of the two haplotypes encoding leucine at codon 198 had 18 sequences, while the other was only sequenced once, suggestive of PCR or sequencing errors. Of the 'susceptible' haplotypes, two were common (23 and 32 sequence copies per haplotype), while the last was only sequenced twice. From

analysing this short 16 bp region within a subset of individuals on this farm it is evident that moderate diversity is present within the population. This diversity was maintained over the course of the season and despite the use of different anthelmintic treatments. Although there are a few common haplotypes these include two commonly sequenced sensitive haplotypes, suggesting the presence of sufficient *refugia* or survival of heterozygotes to maintain such diversity.

### 3.3.3 Assessing population change over time

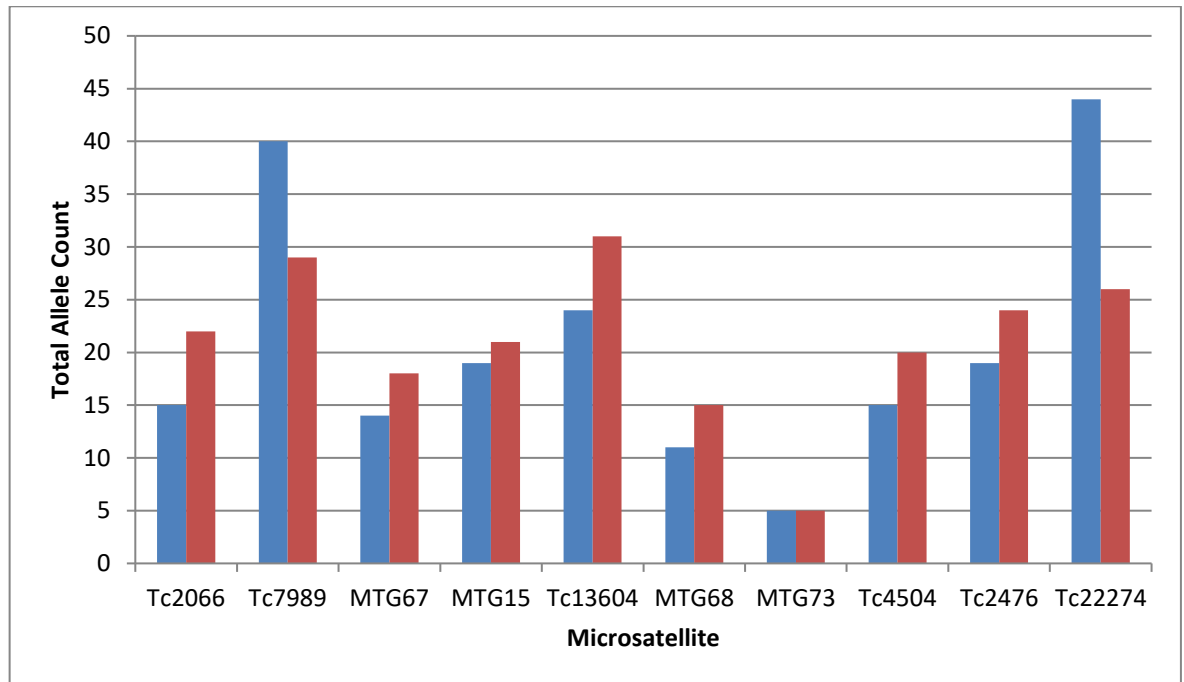
Ten previously published microsatellites (Grillo et al., 2006; Redman et al., 2008) were used to assess the *T. circumcincta* population on the farm over time and in different ages of sheep. At each sampling time point 30 *T. circumcincta* L3 were genotyped and population genetics analysis was conducted using Arlequin v3.5.2.2 (Excoffier et al., 2005). Two ewe samples (collected pre-MOX treatment, and in early June) and several lamb samples (days 0 and 14 of the BZ FECRT and days 0, 7 and 14 of the IVM FECRT) were tested. There was high within-sample diversity, with the majority of variation detected between individual worms rather than between sample groups - only 0.66% of all variation was found to be between samples (Table 3.11).

The total number of alleles in the population per microsatellite varied from five (MTG73) to 44 (Tc22274). Figures were similar to those reported by Redman et al. (2015) (Figure 3.16), suggesting a similarly high diversity on Farm 1 as is present across other UK farm populations of *T. circumcincta*. For many microsatellites, the presence of private alleles unique to a sample population was detected across the samples tested, likely reflecting a small sample size and a highly polymorphic population rather than substructure. For some microsatellites the presence of three alleles amplified from a single individual was detected. These were checked for repeatability and, for that individual, all alleles of that microsatellite were converted to 'nulls' for population genetic analysis. This was because these were (i) uncommon and (ii) it was not possible to determine which two alleles would be haplotypic with other microsatellite alleles which were apparently diploid. As the software required diploid data, to remove one allele without just cause would introduce bias.

Overall, only one microsatellite locus, MTG73, was identified as a significant outlier in  $F_{ST}$  analysis between the seven sample groups tested ( $F_{ST}$  value 0.04,  $p < 0.05$ , Table 3.12, Figure 3.17 A). Nonetheless, when repeating the analysis using sample subsets (either only ewe samples, only BZ FECRT samples or only IVM FECRT samples), no selection at this locus was apparent (Table 3.12). No microsatellite appeared to be under selection by IVM treatment. Some microsatellites appeared genetically different between the ewe or BZ FECRT samples ( $p < 0.05$ , Table 3.12 and Figure 3.17). However, this is potentially due to stochastic sampling of a diverse population. Overall pairwise comparisons between samples also suggested significant differences ( $p < 0.05$ , Table 3.13), but again, no consistency was observed. No obvious selection occurred due to either anthelmintic treatment.

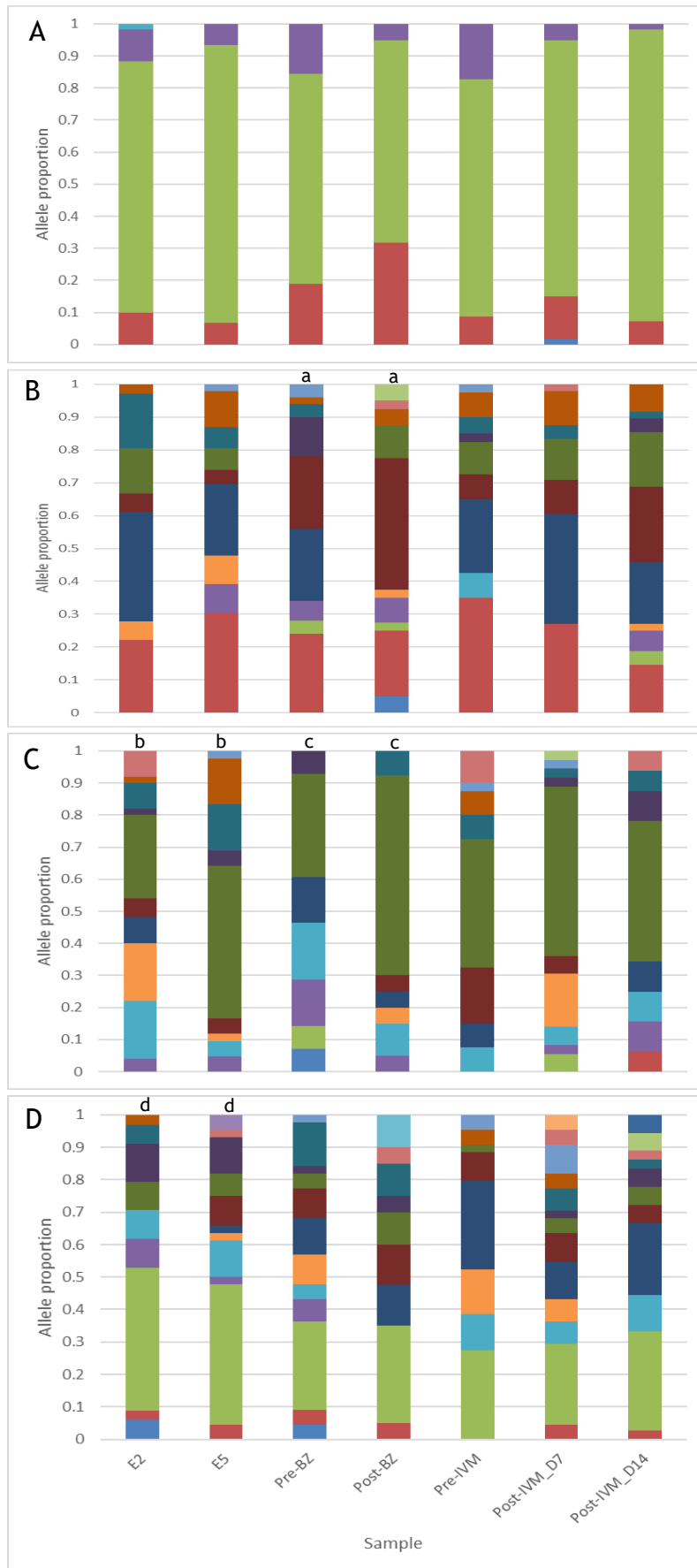
**Table 3.11: Microsatellite global AMOVA results as a weighted average over all loci.**

Source of variation	Variance (%)	Average F statistic over all loci	P-value
Among samples ( $F_{ST}$ )	0.66%	0.00658	1.00
Among individuals within samples ( $F_{IS}$ )	48.73%	0.49052	0.00
Among individuals within total population ( $F_{IT}$ )	50.61%	0.49388	0.00

**Figure 3.16: Microsatellite allele counts comparing Farm 1 data (blue) and data collected from seven UK farms by Redman et al. (2015) (red).****Table 3.12: Individual microsatellite  $F_{ST}$  values between samples.**

Microsatellite	All samples	Ewe samples only	BZ FECRT samples only	IVM FECRT samples only
MTG15	0.0037	-0.0017	0.0062	0.0179
MTG67	0.0008	0.0092	-0.0215	0.0092
MTG68	0.0164	0.0239	0.0024	0.0232
MTG73	<b><u>0.0388</u></b>	-0.0046	0.0081	0.0289
Tc13604	-0.0046	-0.0153	-0.0157	-0.0066
Tc2066	0.0178	-0.0095	<b><u>0.0336</u></b>	-0.0009
Tc22274	0.0036	0.0049	-0.0060	0.0107
Tc2476	-0.0013	<b><u>-0.0269</u></b>	-0.0195	-0.0177
Tc4504	0.0108	<b><u>0.0335</u></b>	<b><u>0.0414</u></b>	0.0015
Tc7989	-0.0002	0.0021	0.0031	-0.0041

Significant allele differences between samples are highlighted in bold and underlined ( $p < 0.05$ ) and may suggest genetically differentiated loci between sample groups. Negative values may be due to rare alleles within a population and should be interpreted as zero.



**Figure 3.17: Sample allele proportions for Farm 1 microsatellites. A. MTG73, B. Tc2066, C. Tc4504, D. Tc2476. a-d denote significant  $F_{ST}$  value by pairwise comparison ( $p < 0.05$ , see Table 3.12). Each colour represents a different allele. Colours should only be interpreted within a plot, not between microsatellites.**

Table 3.13: Sample pairwise  $F_{ST}$  values for Farm 1.

	Ewe April	Ewe June	Pre-BZ day 0	Post-BZ day 14	Pre-IVM day 0	Post-IVM day 7
Ewe April	NA					
Ewe June	0.004180	NA				
Pre-BZ day 0	-0.002450	<b><u>0.018800</u></b>	NA			
Post-BZ day 14	<b><u>0.016270</u></b>	0.011070	0.007600	NA		
Pre-IVM day 0	0.003950	0.015030	0.006410	0.013550	NA	
Post-IVM day 7	0.001710	0.011340	<b><u>0.017300</u></b>	0.016060	0.012650	NA
Post-IVM day 14	0.004250	0.011120	0.007050	<b><u>0.021240</u></b>	0.014350	0.015890

Significant differences between samples are shown in bold and underlined ( $p < 0.05$ ). Negative values should be interpreted as zero.

### 3.4 Discussion

#### 3.4.1 Parasite abundance and diversity on a farm today

Despite the use of multiple new anthelmintics, the frequent use of such anthelmintics, and the presence of anthelmintic resistance in one or more species on this farm, data collected over the year from both ewes and lambs has suggested that the epidemiology of PGE is similar to past descriptions in the UK (Morgan et al., 1951; Boag and Thomas, 1977; Wilson et al., 2008). The peri-parturient rise (PPR) in FEC at lambing was observed in the housed ewes, which was curtailed by an efficacious MOX treatment (Jansen, 1987; Taylor et al., 1997). Following this, a slight rise in egg output was noted over the following few weeks, potentially due to *T. circumcincta*, demonstrating ‘tail resistance’ (Kerboeuf et al., 1995; Sargison et al., 2012), or the presence of other species, identified on this farm, against which MOX has no persistent action. Lamb egg output rose in late August, mirroring the rise in ewe egg output, as expected (Salisbury and Arundel, 1970).

In June, a small increase in FEC in the ewes was noted before rapidly falling, which corresponds with the last recording of a moderately high *T. circumcincta*

prevalence in ewe coprocultures. This phenomenon of an acute change in *T. circumcincta* prevalence has been noted elsewhere (Jansen, 1987; Sargison et al., 2012) and potentially reflects reassertion of immunity against this parasite. That this is delayed until June may suggest that the persistency of MOX prevents resumption of immunity (Brunsdon, 1966). If so, this is a concern for farmers as eggs deposited in June in the UK have a much greater chance of contributing to infective L3 on pasture than those deposited in March (Gibson and Everett, 1971).

The source of infective L3 for lambs arises either from overwintered L3 on pasture or from the ewes, of which the major contribution arises from the PPR phenomenon (Jansen, 1987). In the lambs the FEC remained low throughout the spring and summer. This may be due to the presence of low levels of infective L3 on pasture or the use of neosuppressive BZ treatments in the spring, despite the presence of several BZ resistant species. Grass sward height was short for a considerable period of the early spring and summer, which may have increased the ingestion of infective larvae by ewes (van Dijk and Morgan, 2011), reducing the infective larval burden present for the lambs, especially as egg output from ewes remained low.

The cause for the rise in ewe FEC in November and December (Figure 3.4) does not appear to be due to the presence of a new species, against which immunity needs to be re-established. Neither does it coincide with an increased proportion of *Trichostrongylus* species which are considered to be important in late autumn/winter PGE in the UK (Craig et al., 2006). Hormonal changes induced by the shortening day length or breeding related activity, perhaps leading to reduced food intake or reduced immunity might be partially responsible. Such changes might also be related to climatic differences, perhaps resulting in changes in feed availability and reduced protein intake affecting the immune response (Sakkas et al., 2012). Nevertheless, the grass sward was not noticeably short, especially compared to the summer months when grass was in poor supply due to warm, dry weather and ewes were lactating. An overall increase in the larval availability on pasture may be the cause of this rise, although this is unlikely, as the ewes were grazing pasture which had been used by both ewes



and lambs earlier in the year. If this were the case, one would expect an earlier rise in FEC, coinciding with the rise in lamb FEC in August.

The species diversity on the farm over the season was moderately high, with seven strongyle species (including *N. battus*), from the abomasum, small and large intestines identified. In addition, other internal parasites including *Trichuris ovis*, *Strongyloides papillosus*, *Moniezia* and *Coccidia* were identified. Of the strongyles detected, *T. circumcincta* was the predominating pathogen over the season. The presence of considerable proportions of *C. curticei* was interesting and this small intestinal species has been noted to rise in the early autumn in lambs (Boag and Thomas, 1977). Boag and Thomas (1977) had considered that there was just one cycle of this parasite during the season, however on this farm it is clear that patent infection with *C. curticei* is present in the ewes much earlier in the year, and it is not obvious as to why the presence of *C. curticei* in the lambs is so delayed. The pre-patent period is comparatively short, only 14-15 days (Roeber et al., 2013a), and the development to L3 on pasture takes about 25 days at air temperatures of 13.9 °C (Ahluwalia and Charleston, 1974). Local weather data (Figure 3.18) kindly provided by S. Leeson, (Centre for Ecology and Hydrology, NERC, Figure 3.18) and Cosmos (2018), indicated that favourable temperatures should have allowed eggs deposited in June from the ewes to develop to infective L3 within a month. One study investigating co-infection with *T. circumcincta* and *T. vitrinus* (a small intestinal parasite), noted that *T. circumcincta* infection reduced establishment of *T. vitrinus* (Jackson et al., 1992b). A second study reported reduced establishment of *T. colubriformis* when lambs were co-infected with this parasite and *T. vitrinus*, compared to monospecific infections (Roy et al., 2004). A similar phenomenon may have occurred here. Although considered a mild pathogen by those in the field (Abbott et al., 2012; Roeber et al., 2013a), few studies have investigated its effects on production (Ahluwalia and Charleston, 1975) and its impact as part of a mixed species infection may be different. Given the high prevalence of this species more up-to-date information as to its pathogenicity in the UK may be useful.

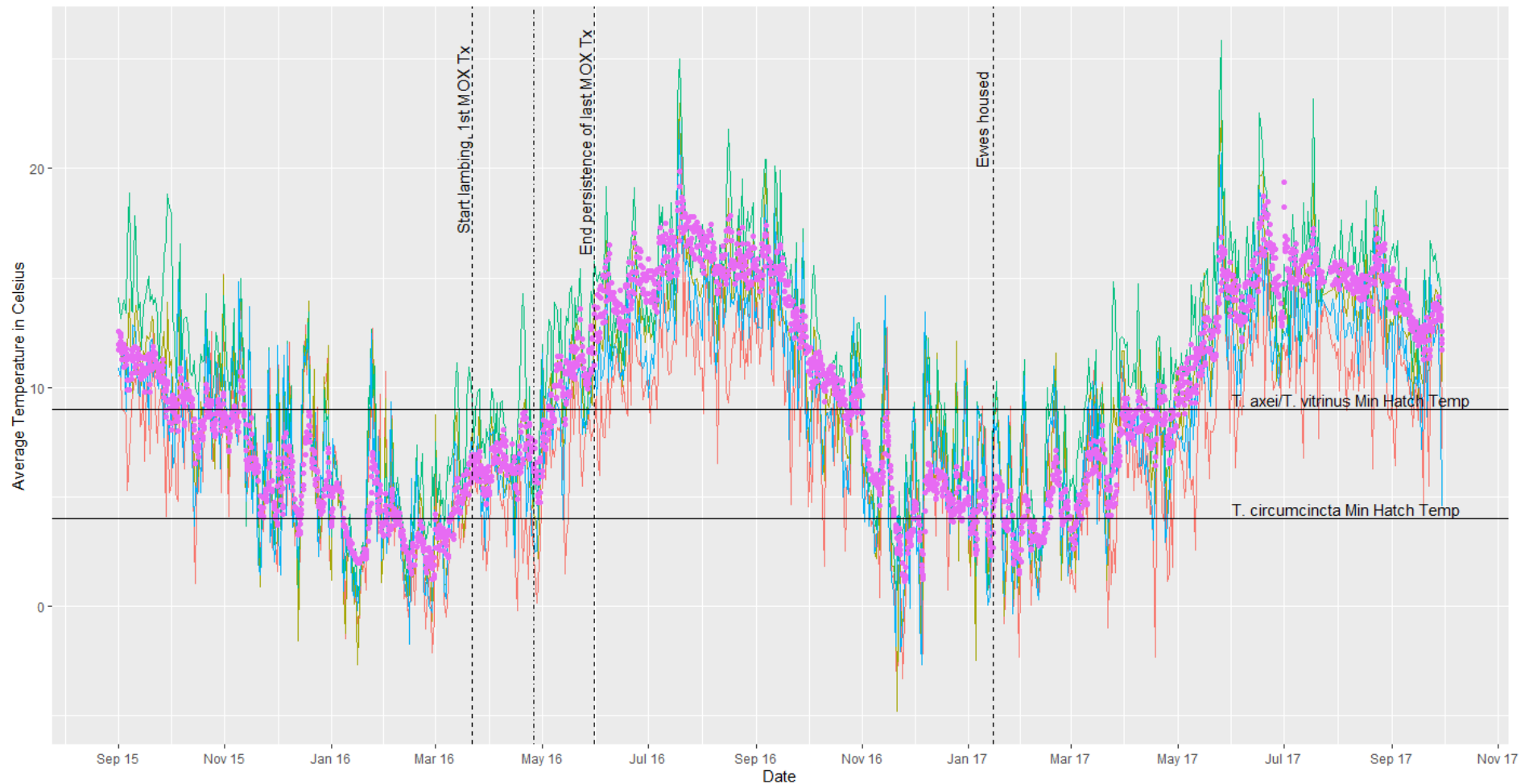


Figure 3.18: Average soil and air temperature between the 1<sup>st</sup> September 2015 and the 30<sup>th</sup> September 2017 collected by S. Leeson (CEH, NERC) (Cosmos, 2018). Purple dots indicate the daily average soil temperature. Coloured lines indicate the daily average air temperature, measured in four time periods throughout the day. The minimum hatching temperature for *Teladorsagia circumcincta*, *Trichostrongylus axei* and *Trichostrongylus vitrinus* is shown (Crofton, 1965).

Similar levels of diversity have been found on other farms in the UK, although far fewer species have been identified on many farms (Burgess et al., 2012; Hybu Cig Cymru, 2015; Redman et al., 2015). Where no anthelmintics are used, greater diversity is found (Craig et al., 2009; Sinclair et al., 2016). Co-grazing untreated animals (e.g. ewes, lambs, gimmers or rams) alongside treated sheep may be key to maintenance of species diversity on this farm. In addition, maintenance of diversity may be due to the presence of multiple BZ resistant species within the strongyle community. Several pasture types were present and these, in addition to the comparatively intensive grazing of pastures may also contribute to species diversity (Silvestre et al., 2000). Silvestre et al. (2000) found that the number of flocks of different origin originally contributing to a farm was an important variable in species diversity. The relevance of this factor on UK farms is hard to quantify. Although many farms, including this farm, are 'openly' managed, with replacement breeding sheep bought each year, new sheep are usually treated with an anthelmintic, reducing the opportunity for introduction of new species onto a farm.

In addition to larvae surviving over winter on pasture, or the presence of such larvae being maintained by winter grazing sheep, it was clear from this study that patent infection with all species was present in the housed ewes at lambing time, indicating overwinter survival in the host (Stadaliènè et al., 2015). Given the absence of *T. circumcincta* in the FEC from housed ewes in January, the presence of *T. circumcincta* egg output at lambing was either from resumption of development of arrested L4 (Reid and Armour, 1975; Smith, 2007), or from resumption of egg laying in adults previously suppressed by the ewe's immune response (Jeffcoate et al., 1992; Stear and Bishop, 1999; Stear et al., 1999). It is clear that immunity *in vivo* does not equally suppress all species, given the difference in species composition of the ewe faecal egg output after June (Morgan et al., 1951). With the continuing rise in anthelmintic resistance (Kaplan and Vidyashankar, 2012), it may be beneficial to better understand the impact of diversity of strongyle infection within a host, how this affects production parameters (Coop et al., 1986; Giudici et al., 1999), and how it can best be managed or exploited for sustainable sheep production in the future. Cross-immunity to different pathogens may be important in natural infections (Reinecke et al., 1980; Coop et al., 1986; Jackson et al., 1992b), and the

maintenance of production, when infected with multiple species, in the absence of anthelmintic treatment may indicate a role for diversity to regulate parasite burdens and egg output (Roy et al., 2004).

Interestingly, there was a marked difference in species composition of the faecal egg output between the ewes and the lambs, particularly throughout July to November. This is an important point to consider when identifying the need for treatment of animals, management of parasites within the flock over the season and interpreting anthelmintic efficacy (McKenna, 1997). The pathogenicity reports of different species are usually based on single species infections and are relative to the overall intensity of infection. However, pathogenicity may also be relative to the age and reproductive status (Taylor and Pearson, 1979b; Jackson et al., 1988), breed (Ross, 1970; Gonzalez et al., 2008) and sex (Craig et al., 2006) of the host. *T. circumcincta* (Coop et al., 1981), *Trichostrongyle* species (Coop et al., 1979; Taylor and Pearson, 1979a; Beveridge et al., 1989), *N. battus* (Crofton and Thomas, 1951) and *H. contortus* (Yadav et al., 1993) are considered highly pathogenic. In contrast, *C. curticei* (Ahluwalia and Charleston, 1975), *C. ovina* and *O. venulosum* (Roeber et al., 2013a) are only mildly pathogenic. These differences mean that species identification for routine FECs could add value in flock health planning on a farm. On this farm, the use of ewes after June to 'hoover up' larvae from pasture could be beneficial. These animals would have a bias against output of *T. circumcincta* eggs onto pasture, whilst increasing the *refugia* of less pathogenic species. This may allow the judicious use of pastures following weaning, particularly if weaning earlier (e.g. at 12 or 14 weeks of age (Laws and Genever, 2015)) such that, dependent on weather and farm species prevalence, fields used by ewes and lambs earlier in the season may become useful for finishing lambs at the end of the season.

Overall, the use of both FEC, management data and strongyle species information allows for a much more informed understanding of the parasite abundance and diversity present on a farm. However a clear caveat is the potential difference between egg output and worm burden (Cabaret et al., 1998), the relationship varying between species. In particular, *T. circumcincta* is known to have a non-linear relationship between the number of adults and the eggs shed per adult female (Bishop and Stear, 2000). Coproculture is a useful

technique, however, culture conditions for each species vary, and the optimum temperature and time taken to develop to L3 can bias results towards particular species (Dobson et al., 1992; O'Connor et al., 2006).

### 3.4.2 Was the anthelmintic use appropriate?

The need to preserve anthelmintics requires that they are only used when necessary. On this farm FECs were rarely greater than 250 epg, which might indicate that treatments were unnecessary (Abbott et al., 2012). Conversely, it is important to realise that these low FECs may have been due to strategic treatments given throughout the year.

The use of MOX to suppress the PPR has been recommended to prevent contamination of the pasture for growing lambs (Salisbury and Arundel, 1970; Taylor et al., 2016), improving lamb growth rates (Coop et al., 1982). For this reason, and to reduce the parasite burden in the ewes, improving lactation (Mavrot et al., 2015), the ewes on this farm were treated with MOX at lambing. The treatment was highly effective at reducing egg output, and egg counts were low in general over the season in ewes and lambs. Had the PPR not been suppressed, daily egg output from a ewe with a mean epg of 600 could have seeded the pasture with approximately 840,000 eggs per day (Waghorn et al., 1999). Thus, this treatment may have been important and had the effect desired. However, the overall benefits of suppressing the PPR have been recently questioned (Taylor et al., 2016; Learmount et al., 2018). It may be that the benefit of treating ewes varies between farming systems and years, dependent on hosts, feed availability, pasture larvae and spring weather. PGE risk versus risk of anthelmintic resistance development should be considered for each farm from year to year.

One downside of MOX use on this farm may be the selection for resistant parasites, at a time when *refugia* is low (Martin et al., 1981). Lambs suckling treated ewes also receive MOX in the milk, exacerbating selection for macrocyclic lactone resistant parasites (Dever and Kahn, 2015). Use of a short acting product would allow curtailment of the PPR, and ewe FECs could be monitored for additional treatments required. On this farm, LEV may be the most suitable anthelmintic, due to the presence of dual IVM and BZ-resistant *T.*

*circumcincta*, although the efficacy of LEV against this parasite specifically should first be tested. Before lambing started the farmer advised that mostly triplets, and only some twins or single bearing ewes would be treated with MOX. However, the medicine book records that 350 ewes received MOX, suggesting that most ewes were treated, leaving few untreated ewes able to seed the pasture with susceptible parasites. Provided that ewes receive adequate nutrition, suppression of the PPR in all ewes should not be necessary, as the pasture larval levels should be controllable with only some ewes treated (Houdijk et al., 2000; Sakkas et al., 2012). On this farm, the grass sward was very short at the start of the season, reflecting the cold winter preceding the start of the study. For this reason, nutrition available to the ewes may have been lower than ideal, leading to increased treatment of the flock to maintain adequate levels of milk production and lamb growth. Nevertheless, it would also be likely that fewer L3 would have survived overwinter on pasture, reducing the need for a persistent anthelmintic.

In the UK, unless it is possible to use a field for young lambs which has not had sheep on it for at least the previous two years, the presence of *N. battus* requires the pre-emptive use of anthelmintics in the spring (Thomas, 1991), based on NADIS parasite forecasts (<https://www.nadis.org.uk/parasite-forecast.aspx>). The use of BZ anthelmintic treatments for *N. battus* are therefore justifiable as this parasite is highly pathogenic in young lambs (Crofton and Thomas, 1951). However, on this farm, anthelmintics were not used until after *N. battus* eggs were detected in the faeces, suggesting that some pathology had already occurred, restricting optimal growth rates over the season. In addition, by the timing of the last dose, lambs may have acquired some immunity towards *N. battus* (Taylor and Thomas, 1986; Israf et al., 1997) and this dose may have been superfluous, serving only to increase selection for BZ resistance. Anthelmintic use at this time in young lambs has the disadvantage of low *refugia* levels, and it is not surprising that several species were resistant to BZ, with BZ resistance associated SNPs almost fixed within the *T. circumcincta* population. The presence of dual-resistant *T. circumcincta* is likely also related in part to this management practice - i.e. BZ selection upon a population already comprised of a significant proportion of parasites either fully or partially resistant to MOX. Following treatment of ewes at lambing, MOX

resistant, and potentially also IVM resistant (Ranjan et al., 2002; Ménez et al., 2016), eggs would be output by the ewes and predominate on pasture, for ingestion by lambs. In addition, lambs could ingest MOX via the ewe's milk (Dever and Kahn, 2015), further selecting for resistant parasites. Therefore the strongyle community within the lambs when they were dosed with BZ would be likely to contain a large proportion of MOX and potentially IVM resistant parasites. Nevertheless, as BZ resistance is common on this farm, and the *N. battus* population appears sensitive to BZ, this was a good choice of anthelmintic. It reserves the use of other anthelmintic classes to treat other species later in the year when they are more significant. In addition, the farmer left 5-10% of the lambs untreated, and ewes were co-grazing with the lambs at this time. These untreated sheep will have continued to contribute to the pasture larval population, and will have been important in maintaining BZ sensitive parasites.

Lamb FECs on this farm were low, and if the guidance threshold of 250 epg for a predominantly *T. circumcincta* infection was used to prompt treatment (Abbott et al., 2012), no anthelmintic treatments would have been provided to the lambs other than for *N. battus*. Although only one other anthelmintic treatment was given, its requirement is hard to justify based on the FEC and speciation data alone. In September, the LEV treatment when the mean FEC was 172 epg, with *T. circumcincta* contributing only 50 epg, was questionable. However, using a different anthelmintic class than was used earlier in the year is recommended (Abbott et al., 2012). Two FECs preceding the treatment revealed a mean epg more than four times as high as the pre-weaning FEC samples. Responding proactively to reduce pasture contamination from a rising FEC was wise, given the high proportions of *T. circumcincta* detected in the lamb samples previously. Hindsight is useful, but in practice welfare and economic concerns require decisions to be made in advance. Lambs in Scotland are fitted with electronic ear tags before they leave the farm, and using these for regular monitoring of lamb growth with an understanding of feed nutritional value, can allow for targeted selective treatment (McBean et al., 2016). This does not require knowledge of the FEC, or species present (assuming other causes of poor weight gain are not a concern), and has been shown to maintain production parameters

(Kenyon et al., 2013; McBean et al., 2016) which may have been of more use on this farm than FECs.

### **3.4.3 Can we accurately and cheaply determine which anthelmintics will be effective?**

Probably the greatest confounding factor limiting interpretation of the FECRT (and the *in vitro* bioassays) was the species diversity. The presence of a highly dual-resistant population of *T. circumcincta* contributing only a minor proportion of eggs pre-treatment, allowed for the calculation of a considerably more effective FECR percentage than was truly observed (McKenna, 1997). The change in species over time on this farm requires a FECRT with species identification to be performed when the farmer would usually treat the sheep - with the age group requiring treatment. The identity of BZ resistant *T. axei* would allow a veterinary surgeon to suggest that BZ anthelmintics are avoided during the late autumn/winter months if 'black scours' was the clinical diagnosis (Craig et al., 2006). However, ideally the FECRT would be repeated when this species was more prevalent. On this farm therefore, reduced anthelmintic efficacy cannot be correctly interpreted without species data. But species identification carries additional costs, and is not routinely performed in the UK. Nevertheless, where an anthelmintic is highly effective, such as in the case of MOX on this farm, the FECRT is a useful confirmation of success. It could also aid diagnosis for causes of weight loss, poor growth and diarrhoea or anaemia not caused by PGE by ruling out anthelmintic resistance, while simultaneously providing beneficial information to the farmer. In addition, although there are limitations to the interpretation of a FECRT as to the number of adults remaining within the host (Cabaret et al., 1998), it is an excellent indicator of the reduction in pasture contamination.

The use of bioassays such as the EHT and LDT provide an understanding of the anthelmintic resistance phenotype of free-living stages of strongyles. Although these may not necessarily reflect the *in vivo* mechanisms of resistance (Rufener et al., 2009a; Barrère et al., 2012; Kotze et al., 2012), they are quicker than performing a FECRT, identifying anthelmintic efficacy within just 48 hours to a week. Bioassays only require a single faecal sample collection, and do not contribute to drug selection on farm. In this study, the EHT worked



comparatively well and identified the presence of dual-resistant parasites without the need to treat hosts with two anthelmintics. Over the season, the percentage hatch in the definitive dose well reflected recent BZ anthelmintic treatments, reducing as time since the last treatment increased. Variability between replicate wells of a test increased as the species diversity increased, and the abundance of *T. circumcincta* reduced. An exception was the test performed at the end of August, when there was high variability between wells, despite a high percentage of *T. circumcincta*, suggesting that there may have been some variation in the resistance phenotype of the *T. circumcincta* population at that time. However, without PCR speciation of eggs and L1 larvae, which is time-consuming and can be technically challenging post EHT, the results are difficult to interpret. For both tests, as for coproculture, it is important to remember that not all species will develop equally well at a set temperature, and over a set time period (Dobson et al., 1992). This was potentially observed in the present study, with differences in the control wells pre- and post-treatment. This may bias results towards those species more suited to the test set-up, compromising interpretation of the anthelmintic efficacy.

While the EHT appeared to detect BZ resistance reasonably well, and has been ring-tested in Europe using lab isolates of *H. contortus* (von Samson-Himmelstjerna et al., 2009a), the LDT did not prove as useful in this study. There was little change in the ED<sub>50</sub> post-IVM treatment in September (RR = 1.7), despite a considerable reduction in the proportion of IVM sensitive species *in vivo*. Similar RRs were also identified between IVM resistant and susceptible isolates of *H. contortus* using an agar based LDT, when IVM was the anthelmintic tested (Gill et al., 1995). However, larger RRs were obtained using more polar avermectins, and using a different avermectin as a proxy measure of IVM resistance may provide more informative measures of IVM phenotype (Gill et al., 1995). In contrast, there was a marked difference between the pre- and post-BZ treatment populations ED<sub>50</sub> in the EHT (RR = 12.8). In addition, although the EHT pre-treatment suggested a generally susceptible population, there was the presence of a more resistant sub-population detectable by hatching percentages at the 0.2 and 0.3 mg/ml concentrations (von Samson-Himmelstjerna et al., 2009a). Collectively these results may suggest a similar genetic basis for BZ resistance both *in vivo* and *in vitro* in this field population. In contrast, different

genetic mechanisms may underlie *in vitro* and *in vivo* IVM resistance, or the *in vitro* tests may be inappropriate (George et al., 2017).

Identifying resistance using molecular methods allows for rapid and robust understanding of the resistance status of a population. It can also be performed using eggs, greatly increasing the speed of diagnosis, provided the method is sensitive to extremely low quantities of DNA. For BZ, the presence of the  $\beta$ -tubulin isotype-1 SNPs at codons 167, 198 and 200 have been closely associated with BZ resistance (Kwa et al., 1994; Silvestre and Cabaret, 2002; Ghisi et al., 2007). While these SNPs can be detected in a laboratory setting, no commercial test has been developed for use on farm. In this study, we used pyrosequencing, a technique which provides genotype and some haplotype data for each individual sequenced. The results from this farm clearly indicate the importance of testing all three codon positions. Codon 200 was the most commonly mutated, as has been found elsewhere (Redman et al., 2015; Ramünke et al., 2016). However, in some studies, codon 198 was found to have a higher frequency of mutations than codon 200 within the sample population tested (Redman et al., 2015; Esteban-Ballesteros et al., 2017). Most studies have associated leucine at P198 with BZ resistance in *T. circumcincta*, as was found here, but some have found alanine (Esteban-Ballesteros et al., 2017), which is more commonly identified in *H. contortus* (Chaudhry et al., 2015), or valine (Avramenko et al., 2019). Although it may be a sequencing error (n=1), the presence of a stop codon was detected at codon 198 in this study. This diversity increases both the cost and the time taken to obtain a conclusive answer. The need to first select and speciate individuals restricts the use of molecular techniques. Newer techniques, including MiSeq, are now being used to sequence pools of strongyles to determine BZ resistance haplotypes, which may be extremely useful in the future (Avramenko et al., 2019).

Results from this PhD and a recent study in Ireland (Keegan et al., 2017a), indicate that BZ resistance in *T. circumcincta* may be more complicated than just the presence of SNPs at these three codons, and that the presence of resistant individuals may be missed by molecular methods focusing solely on these. In this study, the presence of homozygous 'susceptible' offspring indicates the survival of adults hetero- or homozygous for BZ sensitive alleles. In

Ireland, Keegan et al. (2017a) identified the presence of adults surviving oxfendazole treatment, which were genotyped as heterozygous or homozygous 'susceptible' at each of the BZ SNPs. Similar ideas of alternative BZ resistance mechanisms have been suggested previously. Originally, selection at both  $\beta$ -tubulin isotype-1 and  $\beta$ -tubulin isotype-2 was noted in BZ resistant isolates compared to susceptible isolates (Beech et al., 1994; Lubega et al., 1994). It was suggested that BZ resistance could occur following mutation in  $\beta$ -tubulin isotype-1 but that high level phenotypic resistance was possible with deletion of  $\beta$ -tubulin isotype-2 in *H. contortus* (Kwa et al., 1993). In *Caenorhabditis elegans*, multiple *ben-1* gene variants have been associated with BZ resistant phenotypes, and the N2 Bristol strain, commonly used in research, was unusual in having the codon 200 SNP (Driscoll et al., 1989; Hahnel et al., 2018). Similar to *H. contortus*, highly BZ resistant *C. elegans* strains had a deletion of *ben-1* (Driscoll et al., 1989). Resistance phenotype may also vary by codon (Silvestre and Cabaret, 2002) and lifecycle stage (Barrère et al., 2012; Kotze et al., 2012). In addition to the target gene(s), studies have reported other mechanisms which may increase BZ resistance or tolerance, including the multidrug transporters, the P-glycoproteins (Blackhall et al., 2008). It is clear from EHTs that there are a range of resistance phenotypes even within a single isolate population, with only a subset of worms able to survive at higher concentrations. This would suggest a range of genotypes underlying this trend, which may be more relevant in some field populations than others.

Although molecular methods can be extremely useful and reliable, despite the potential for multiple mechanisms of resistance, they are comparatively expensive. For example, taking the Animal and Plant Health Agency (APHA, a standard lab for diagnostic testing in the UK) as a guide, a FECRT would cost a farmer £25.20 for a McMaster FEC for a single sheep (pre- and post-treatment), providing five or more animals were included. A parasitological examination of a single faecal sample would be £34.30. Morphological identification of L3 was priced at £172.30 (price data accessed July 2019). No EHT or LDT was included in the price list. In this study, the approximate cost of speciating 96 individual strongyles to identify four species by a multiplex PCR was £37.25, not including staff costs. The pyrosequencing assay, not including speciation prior to genotyping, was approximately £240 per 30 worms genotyped at all three  $\beta$ -

tubulin isotype-1 SNPs. Although these costs would be offset by the size of the flock, and therefore the cost-benefit spread over many individual sheep, for many farmers the cost would be likely prohibitive, especially in the face of known resistance. As such, any molecular tests used in research settings need first to be adapted for commercial use, so that either by scale of use, or a cost-reduction in materials, such tests become cost-effective for farmers. For BZ resistance in *T. circumcincta*, which is widespread, the uptake of molecular testing is unlikely. However, if resistance was not yet identified, or there was potential to monitor development of resistance with different management practices, the uptake of such tests might be more likely, for example for emerging resistance of *N. battus* (Morrison et al., 2014) or cattle parasites (Winterrowd et al., 2003), or for IVM resistance.

As both the pyrosequencing and microsatellite results showed, the *T. circumcincta* population on this farm was highly diverse and did not appear to segregate into sub-populations at different times of year, in different age groups of sheep or following either BZ or IVM treatment in September. On this farm the presence of multiple genetic mutations associated with BZ resistance, and multiple haplotypes carrying these alleles, suggests that a soft sweep has occurred as observed elsewhere in the UK (Redman et al., 2015). This may be related to multiple origins of resistance within the farm, for example several resistant haplotypes may have been present since the establishment of the farm. Alternatively, the habit of replacing stock each year may have sequentially introduced new resistance alleles into the *T. circumcincta* population (Redman et al., 2015). In the UK, the stratified flock system, with many farms buying and selling stock at various times of year, may help to maintain species diversity but also increase the chance of resistance being introduced onto a farm. In France, where many farms are established and remain 'closed' with no new stock bought in, resistance was found to be linked to the number of farms of origin at establishment (Silvestre et al., 2000). Fewer  $\beta$ -tubulin isotype-1 haplotypes were identified in French *T. circumcincta* populations compared with UK populations (Skuce et al., 2010). Due to the practice of quarantine drenching (a necessity to reduce the spread of disease and resistant parasites), the likelihood of introducing new sensitive haplotypes, with potentially higher fitness than resistant parasites (if such fitness costs exist), is unlikely. In the presence of

such diversity, studies investigating single genes can also increase the chance of false positive association of a SNP with a resistance phenotype (Laing et al., 2016; Rezansoff et al., 2016), and this may be why there is so much contradictory evidence regarding the mechanism of IVM resistance.

### **3.5 Conclusion**

Methods used to detect anthelmintic resistance on Farm 1 were confounded by species diversity. Molecular tests were useful and revealed high individual diversity within the *T. circumcincta* population, and the presence of multiple haplotypes carrying BZ resistance associated SNP alleles. Although the EHT correlated with BZ resistance phenotype, the LDT was less informative. The genetic basis of IVM resistance is not known, and no molecular test is available. The following three chapters present results of a genome wide association study using two NGS technologies to identify genetic markers of IVM resistance. L3 progeny of *T. circumcincta* field populations, selected by FECRT, are used.

## 4 One site does not fit all: difficulties in identifying markers of IVM resistance using a reduced representation sequencing method

### 4.1 Abstract

Identification of IVM resistance in field populations is currently performed using a faecal egg count reduction test. This test has multiple limitations and can be insensitive to diagnose resistance. In contrast, molecular tests of resistance are sensitive, specific and results can be obtained rapidly. Currently the genetic basis of IVM resistance in *Teladorsagia circumcincta* - the main pathogen of growing lambs in the UK - is unknown and no genetic test exists. This limits research into the effect of management and *refugia* on IVM resistance spread and development. To date, studies investigating single genes or gene families have not identified a conclusive genetic basis for IVM resistance in *T.*

*circumcincta*. Genome wide association studies provide an unbiased approach to identify novel loci associated with anthelmintic resistance. The published *T. circumcincta* reference genome is large (~700 Mb) and highly fragmented (81,000 contigs). Instead, a draft reference genome has been recently assembled (685 Mb, 8025 contigs) and was used for this study. I used a reduced representation next generation sequencing method - double digest restriction-site associated DNA sequencing - to sequence 30 pre-treatment and 32 post-IVM treatment *T. circumcincta* individuals. These were offspring collected from adults treated with IVM. In total 15,912 polymorphic RADloci were used for analysis. My results identified multiple loci that were differentiated between the pre- and post-treatment populations. Nevertheless, linkage decay was rapid and RADloci were unevenly distributed across the genome. Manual annotation of regions up to 75 kb distant of highly differentiated RADloci using RNA-Seq from a published study was performed. A RADlocus containing two of the highest  $F_{ST}$  value SNPs, with apparent selection of the SNP alleles by IVM treatment was identified. Yet closer inspection failed to confirm selection by IVM over this RADlocus. No IVM candidate genes were present on contigs containing highly differentiated RADloci. Individual diversity and the reference genome used may have limited the usefulness of ddRAD-Seq in this field population.

## 4.2 Introduction

Previous studies investigating anthelmintic resistance have mostly focused on a small number of candidate genes (Walker et al., 2001; Dicker et al., 2011b; Martinez-Valladares et al., 2012b). While this has proven remarkably successful for BZ resistance (Kwa et al., 1995), this approach has not been effective for IVM resistance, with studies providing conflicting conclusions on the genetic basis of IVM resistance (Dent et al., 2000; Njue and Prichard, 2004; Martinez-Valladares et al., 2012b; Rezansoff et al., 2016). However, a single microsatellite locus in *H. contortus* (Hc\_8a20) has been repeatedly linked to IVM resistance (Redman et al., 2012; Rezansoff et al., 2016). Recently, with the increasing availability and decreasing cost of next generation sequencing (NGS) methods, genome wide association studies (GWAS) have been undertaken to identify the basis of IVM resistance in parasitic nematodes (Choi et al., 2017; Doyle et al., 2017; Doyle et al., 2019). In contrast to candidate gene studies, these GWAS are both unbiased, allowing novel genetic mechanisms to be identified, and have the ability to test multiple resistance loci in parallel. This latter aspect is especially important as IVM resistance is known to be polygenic in *C. elegans* (Dent et al., 2000; James and Davey, 2009) and more than one mechanism may exist in parasites (Le Jambre et al., 2000; Kotze et al., 2014). The size of parasitic nematode genomes still makes the use of whole genome re-sequencing (WGS) techniques comparatively expensive, however, and reduced-representation approaches - sequencing less than one to five percent of the genome (Altshuler et al., 2000; Peterson et al., 2012) - are particularly attractive in these organisms (Baird et al., 2008).

In 2000, a letter to Nature described a reduced-representation shotgun sequencing method (RSS) for SNP analysis of the human genome. This involved cloning and sequencing fragments of genomic DNA from a pool of ten individuals after restriction digestion and size selection (Altshuler et al., 2000). The use of such a method increased the chance of detecting SNPs between individuals, at a fraction of the cost of WGS and was welcomed at a time when the human genome was still being assembled. Since then, various methods have been developed, including restriction-site associated DNA sequencing (RAD-Seq) (Baird et al., 2008). Originally conceived as a microarray method (Miller et al., 2007), the incorporation of Illumina sequencing greatly increased the scope of the

method (Baird et al., 2008). RAD-Seq involved digestion of genomic DNA using a single restriction endonuclease, ligation of P1 adaptors (modified Illumina adaptors) to the enzyme cut-site and random shearing of the DNA. Following this, a second, divergent 'Y' adaptor (P2) was ligated to the sheared end of the DNA fragment. PCR amplification followed, allowing only those fragments with a P1 adaptor to be amplified. The region immediately adjacent to the P1 adaptor was sequenced, i.e. immediately adjacent to the restriction enzyme cut-site, and the sequence was denoted a 'RAD-Tag'. SNPs within these loci were called RAD markers (Miller et al., 2007; Baird et al., 2008). While this method greatly increased the number of markers it was possible to sequence, and reduced the bias associated with microarray methods, library preparation costs limited the number of individuals that could be multiplexed in a single Illumina lane. Sequencing of loci in both directions from a cut-site also conferred some level of redundancy to the method. Furthermore, without the presence of a reference genome, many sequences were discarded as complexity could not be sufficiently reduced to enable robust SNP calling over sequencing errors (Peterson et al., 2012). The use of restriction endonucleases, specific to a particular pattern within the DNA, by necessity requires all individuals to be homogenous at the cut-sites in order to retain the same DNA fragments for sequencing (Altshuler et al., 2000). Obviously, in nature, this is not the case, and increasing diversity within a population leads to increasing levels of polymorphic cut-sites, such that some RAD-Tags will be lost from each individual within a sample group (Rubin et al., 2012). While this can be used as a test of SNPs within the genome itself (Miller et al., 2007; DaCosta and Sorenson, 2016), it can also lead to heterozygous loci appearing to be homozygous, due to the loss of one haplotype.

Peterson et al. (2012) modified the method to include two restriction endonucleases, a technique known as 'double digest RAD-Seq' (ddRAD-Seq). In this method, the P1 adaptor contained an inline barcode (molecular identifier), to increase the number of samples that could be multiplexed, therefore reducing library set-up costs. The P2 'Y' adaptor contained a second Illumina primer site, allowing paired-end sequencing. ddRAD-Seq facilitated multiplexing of up to 576 individuals within an Illumina flowcell lane. Furthermore, Peterson et al. removed the random shearing step, but included a size selection step. These changes increased both the flexibility of the method (increasing the potential for



fine-tuning the number of RAD-Tags likely to be sequenced), and the specificity of selected RAD-Tags between different individuals, increasing the chance that the same homologous genomic regions would be retained and amplified in each individual, improving the method's efficiency per megabase of sequencing cost (Peterson et al., 2012). Additional benefits included the ability to increase the number of RAD-Tags retained, as additional complexity could be accounted for due to paired-end reads, and the presence of two different restriction cut-sites on each fragment. This also made it easier to perform SNP analysis without a reference genome, and Peterson et al. (2012) developed the means to analyse their ddRAD-Seq RAD-Tags *de novo*. Paired-end reads were of further benefit as they enabled greater confidence in calling variant sites if the reads overlapped. As a sequencing reaction progresses, sequencing errors increase, such that the bases called at the 3' end of a read have a higher chance of being falsely identified than at the 5' end. This is in part due to degradation of the DNA over time. By judicious use of size selection methods, DNA fragments can be chosen so that many will have overlapping paired-end reads. This will increase the sequencing depth at the 3' end of the reads, providing two separate opportunities for sequencing bases at these ends, and improving genotyping confidence (Andrews et al., 2016).

While the budget usually dictates the amount of sequencing possible, experimental design for ddRAD-Seq ideally requires knowledge of the genome size and the frequency of cut sites for any given restriction endonuclease in the organism under investigation. Further to this would be knowledge of the mean and range of fragment sizes produced by the digest (i.e. the number of RAD-Tags expected). This dictates the number of individuals that can be sequenced per lane to achieve the required sequencing depth. However, one of the great attractions of RAD-Seq is the use of the technique in non-model species for which there is not yet a full, or indeed, any reference genome (Hohenlohe et al., 2012). The published *T. circumcincta* genome (PRJNA72569), available during the design stage of this experiment, is highly fragmented (81,000 contigs), and incomplete. Estimates of the number of times a restriction endonuclease might cut could be extrapolated from a similar species, or estimated from previously sequenced genes or regions based on the GC percentage, assuming a uniform GC percentage across the genome (Baird et al.,

2008). This requires an assumption of a similar genome length and frequency of cutting, and this may not be as expected. For example, *T. circumcincta* has an estimated genome size of close to 700 Mb (Choi et al., 2017), yet the closely related parasitic nematode *H. contortus* has a genome size of just 280 Mb (Doyle et al., 2018).

Attempting to understand the effect of population diversity on a RAD-Seq experiment can be complex to predict. The impact of diversity will also be dependent on the goals of the study. Low diversity will impact phylogenetic studies by requiring increased numbers of markers to obtain sufficient diverse markers to separate individuals and traits (Hohenlohe et al., 2010). In contrast, higher diversity populations may lose markers due to polymorphisms in the cut-site (Rubin et al., 2012), but the RAD-Tags remaining are likely to have more informative SNPs (Paris et al., 2017), so fewer markers are required. For GWAS, where RAD markers are required to be linked with a causal SNP, the density of markers may be important. In outbred, diverse populations, such as the nematode *Caenorhabditis remanei*, linkage disequilibrium (LD) may be low, requiring a large number of RAD-Tags to sufficiently capture markers under selection (Cutter et al., 2006). The number of individuals and markers required for a study to achieve sufficient power, when nucleotide diversity and (LD) is unknown, is thus very hard to accurately predict, and is often curtailed by the budget available for sequencing. Despite this, many RAD-Seq studies have been successfully completed (reviewed by McKinney et al. (2017)). Paired-end sequencing alongside comparatively low error rates in Illumina sequencing, and the application of Bayesian based models, permits variant calling from as low as 3 to 10x coverage (Peterson et al., 2012; Maruki and Lynch, 2017). This enables sequencing resources to be spent on increasing the number of individuals and sites sequenced, rather than increasing sequencing depth.

The aim of this study was to identify genetic markers of IVM resistance in UK field populations of *T. circumcincta*. To allow for an unbiased approach, with the potential to detect novel regions of the genome under selection by IVM, a GWAS was performed. As *T. circumcincta* is estimated to have a large genome size, the reduced representation method, ddRAD-Seq, was chosen for the study. Results describe the validation and outcome of this approach to detect genetic

differentiation in progeny of a *T. circumcincta* field population, pre- and post-IVM treatment, from a commercial farm.

## 4.3 Results

### 4.3.1 Farm information and FECRT data

A single UK farm was chosen for the analysis. Farm 2 was located in southwest England, and had previously diagnosed multi-drug resistance (BZ 83.7%, LEV 92.5% and IVM 93.0% efficacy), detected by FECRT in 2013 (Glover et al., 2017). In August 2016 a further FECRT was performed by the farm's veterinary surgeons using four groups of 20 lambs for each anthelmintic tested. BZ, LEV, IVM and MOX efficacies were calculated to be 87.7%, 95.4%, 87.4% and 98.6% respectively (Table 4.1). Little is known about how the farm is managed, however it is understood that two separate flocks were merged between 2013 and 2016. In 2016, pre-treatment samples from all groups of lambs were combined for coproculture and post-treatment samples for each anthelmintic treatment group were coprocultured separately. From the 2016 samples, L3 strongyles were speciated by PCR using the ITS2 region, as described in 2.4.2. Of these, 69.8% were identified as *T. circumcincta* pre-treatment, and only *T. circumcincta* was identified post-IVM treatment. Extrapolation of the strongyle species percentages to the faecal egg count (FEC) data suggested that IVM was 81.9% effective against the *T. circumcincta* population on Farm 2 in August 2016 (Table 4.1, Figure 4.1).

**Table 4.1: Faecal egg count reduction test results on Farm 2.**

Anthelmintic tested	FECRT 2013 (%)	FECRT 2016 (%)	FECRT 2016 against <i>Teladorsagia circumcincta</i> (%)
BZ	83.7%	87.7%	ND
LEV	92.5%	95.4%	ND
IVM	93.0%	87.4%	81.9%
MOX	ND	98.6%	ND

Overall percentage reduction in strongyle egg output following anthelmintic treatment is shown. The proportional reduction in *Teladorsagia circumcincta* faecal egg count is indicated for those samples where L3 larvae were speciated by PCR of the ITS2 region. ND = Not done.

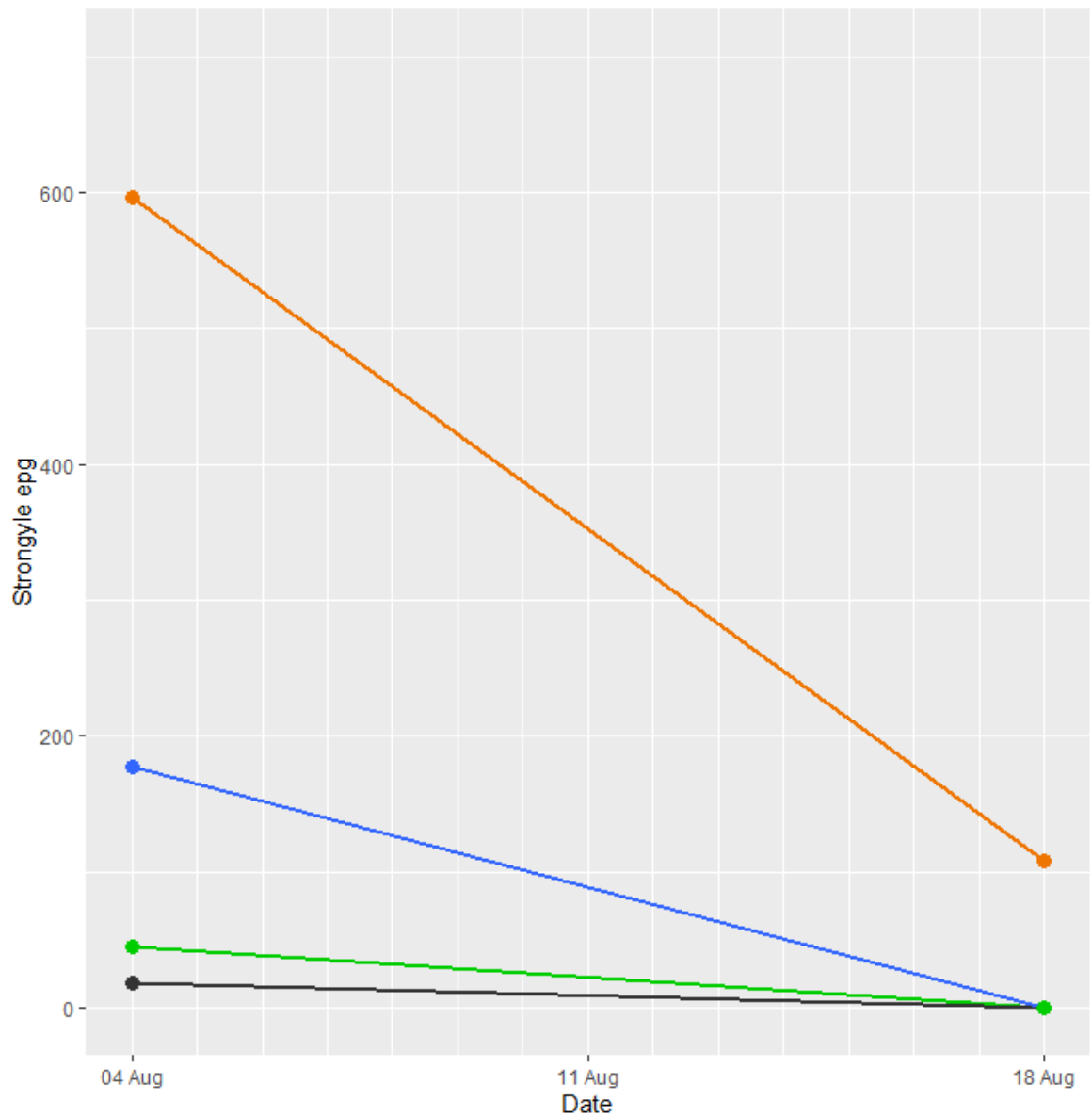
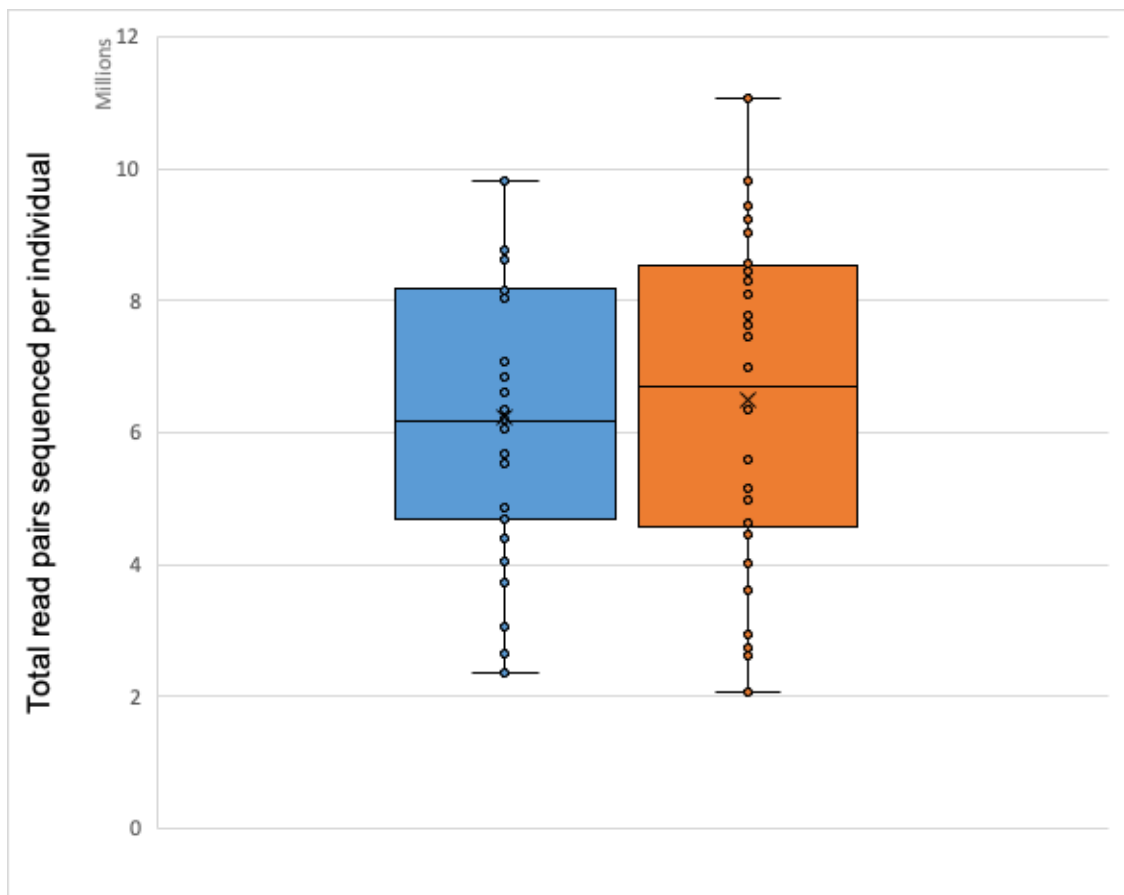


Figure 4.1: Ivermectin Faecal egg count reduction test Farm 2, 2016, with eggs per gram attributed by species. Following coproculture, 96 pre- and 89 post-treatment L3 strongyles were speciated by PCR using the ITS2 region. Species percentages were used to extrapolate eggs per gram of faeces by each species present. Key to colours: green = *Oesophagostomum venulosum*, black = *Trichostrongylus axei*, orange = *Teladorsagia circumcincta*, blue = *Trichostrongylus vitrinus*.

### 4.3.2 ddRAD-Seq read quantity and quality

Two pools each of 31 whole genome amplified (WGA) and barcoded individual *T. circumcincta* L3 (each with 15 pre-treatment and 16 post-IVM treatment individuals to minimise batch effects) were sent to the WSI. Pools were successfully sequenced using 125 bp paired-end reads. Each pool was run in a separate flowcell sequencing lane and the quantity, quality and composition of reads was similar between lanes. Following de-multiplexing, 96.9 and 97.6% of reads were retained per pool.

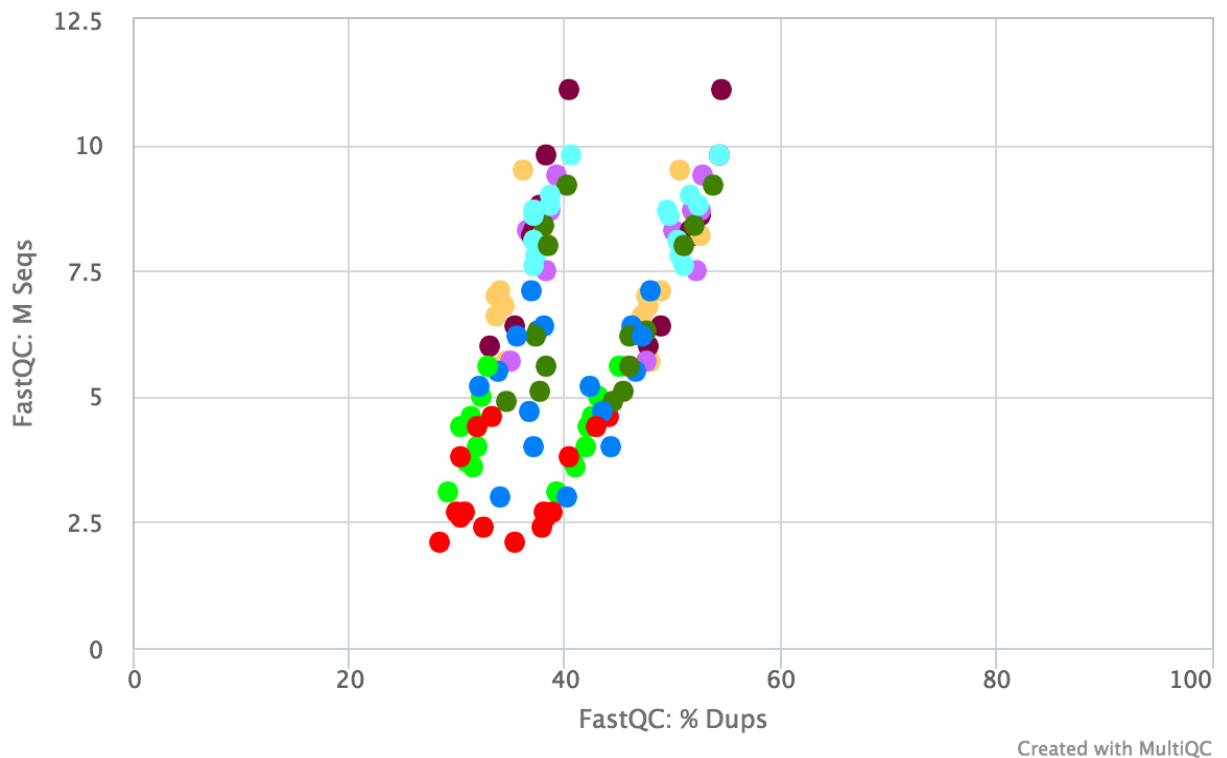
Between 2.1 M and 11.1 M total read-pairs (a total of 4.2 to 22.2 M reads) were sequenced for each individual, with a mean of 6.4 M and a standard deviation of 2.2 M (Figure 4.2, Appendix 3). Values were similar across pre- and post-treatment samples. A high level of read duplication averaging 41% (28.3 to 54.5%) was seen across all individuals. This was as expected for RAD-Seq, due to the same regions being sequenced many times, in contrast to WGS, when lower levels of duplication would be expected. There was a positive correlation between the total number of reads sequenced and the percentage duplication. Percentage duplication was always lower for the second read of a pair, likely reflecting DNA degradation during sequencing i.e. although duplicates (including PCR duplicates) will have been the same (as the reads are in a pair), the second read of each pair will have increased errors, reducing the detection of such duplicates.



**Figure 4.2: Total number of read pairs sequenced per individual. Blue = Pre-treatment, Orange = Post-ivermectin. Each individual is represented by a dot. The mean values for each sample population are shown by an X.**

Overall quality was good, with all read base positions attaining an average base quality Phred score of greater than 30 for each individual, and most positions within reads had a Phred score of greater than 35. A score of 30 equates to a 99.9% probability that the base is correct. Restriction enzyme cut-sites were clearly represented, and base content appeared balanced along the remaining read length. Per base N content was very low, being less than 0.3% for all reads. The GC content was normally distributed for all sets of reads, being slightly higher (46-47%) for reads with the CATG cut-site, as opposed to the reads generated from the AATT cut-site (44-45%).

Samples had been processed in the lab in batches of eight (see 2.5.1). If there was no batch effect, one would expect there to be no relationship between individuals within a batch. However, it is clear that some bias has occurred between groups. The earliest processed WGA samples (Figure 4.3, red and light green) tended to have fewer reads sequenced, and displayed a reduced variation between individuals than some of the later processed samples (dark green, maroon), which had increased reads sequenced per individual, with greater variation between individuals (Figure 4.3).



**Figure 4.3: Identification of sequenced samples by whole genome amplification batch number.** Using MultiQC (Ewels et al., 2016), samples were identified by their WGA batch and the number of sequenced reads (in millions) were plotted against the percentage duplication. Reads to the right are those of the primary read in each pair, while those to the left are the second read in each pair. Batches were processed in the following order: red, light green, dark blue, light orange, light blue, lilac, dark green, maroon. Each point represents one individual.

### 4.3.3 Reads aligned well to the reference genome

Using BWA-MEM (Li, 2013), de-multiplexed reads were aligned to the draft *T. circumcincta* Tci2 genome. This genome is 685 Mb in length, with 8025 contigs, and an N50 value of 288.5 kb, with 642 contigs longer than the N50. The longest contig is 2.08 Mb long. Although a draft version, this is a considerable improvement on the published genome available at WormBase ParaSite, PRJNA72569 (Choi et al., 2017). Of the total number of sequences, on average 82.6% of the reads mapped successfully as a pair (range 76.5 to 85.2%, Figure 4.4). However, fewer reads mapped successfully to the genome in their pair together (denoted ‘properly paired’, 60.6%, range 52.1 to 70.0%). Most of the discrepancy between these figures is due to reads aligning to separate contigs, reflecting the fragmented nature of the genome (Figure 4.4). Secondary alignments ranged from 4.6 to 10.4% (average 7.9%).

The mapping quality (MAPQ) reflects the probability that a read is incorrectly mapped to the genome. Calculated using  $-10[\log_{10}(\text{probability incorrectly mapped})]$ , values of 20 suggest that there is 99% chance the read is correctly mapped. A MAPQ=0 is used to denote a read which maps equally well to two or more genomic positions, but it may align well, or poorly, to those positions. The percentage of total sequences mapping with a MAPQ=0 was on average 13.9% (range 12.4 to 15.3%, Figure 4.4). The average base quality per individual was 35.8 (range 35.6 to 36). Average insert size per individual (overall size, in bp, of both reads of a pair mapped to the genome, including the gap between reads), was 220.43 bp (range 215.2 to 225.3). Noting that ligated adaptors will have been removed before alignment to the genome this insert size reflected the 225-325 bp size selection step using the Pippin Prep (Sage Science). For all mapping statistics there was no significant difference between the pre-treatment and post-IVM sample populations, other than the percentage of total sequences mapped as a pair to the genome ( $p < 0.001$ ). For this statistic, the average for each of the pre- and post-treatment samples was 83.7 and 81.6% respectively. The average error rate (the number of mismatches per bases mapped) was  $3.44\text{E-}02$  (range  $3.34\text{E-}02$  to  $3.53\text{E-}03$ ).



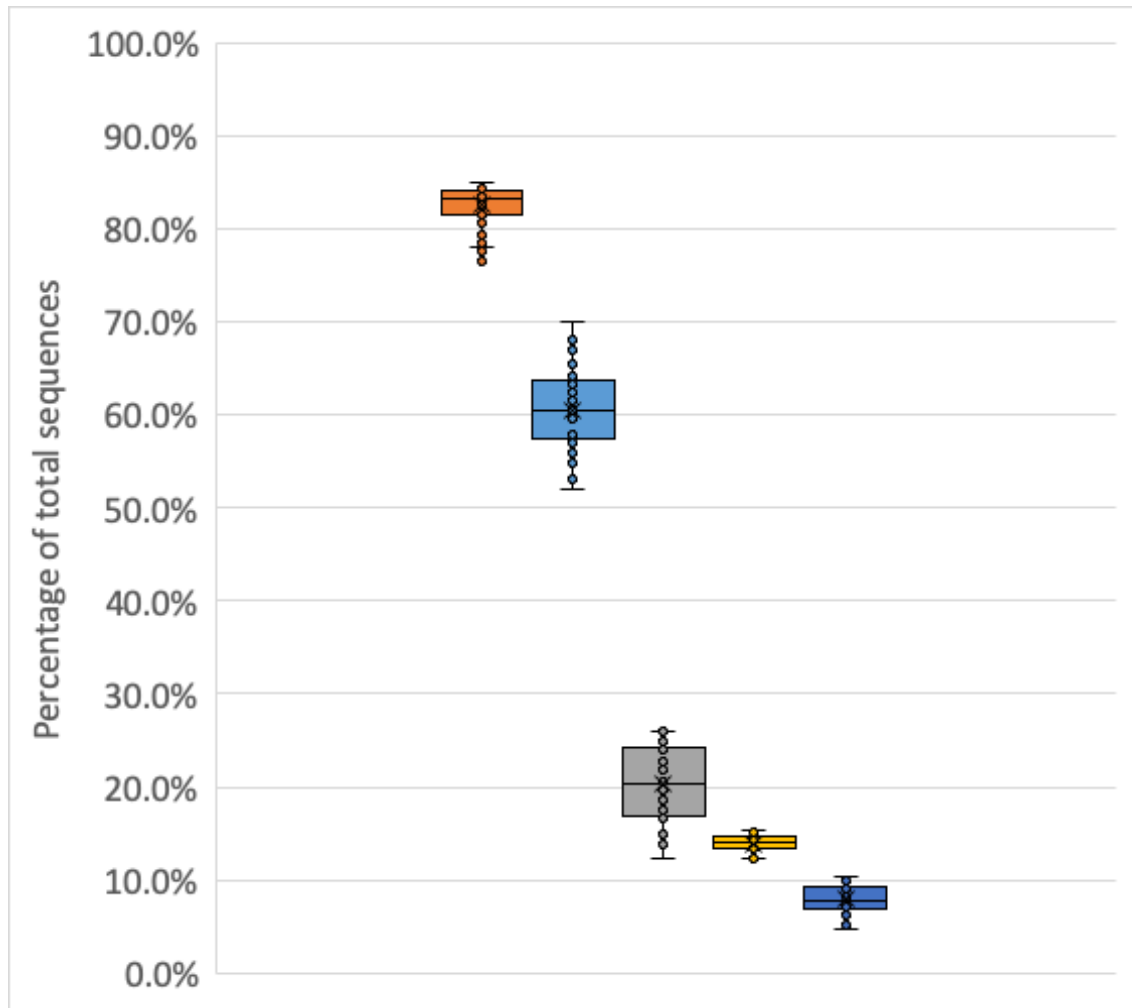


Figure 4.4: BWA-MEM alignment percentages per individual. Each value is the percentage of reads mapped to the reference genome out of the total number of reads sequenced. 'X' is the mean. From the left; orange = 'Mapped as a pair', light blue = 'Properly paired', grey = 'Mate mapped to another contig', yellow = 'MAPQ=0', dark blue = 'Secondary alignments'.

#### 4.3.4 Over 3 million RADloci identified, but very few shared between individuals

In total, 3,341,180 RADloci were identified. Of these, 75,517 RADloci were sequenced in 80% or more of each of the pre- and post-treatment samples (Paris et al., 2017), sampling less than 0.03% of the genome. Most RADloci retained were non-polymorphic (78.9%, 59,605/75,517), with the remainder containing 119,892 variant sites. The 15,912 polymorphic RADloci sampled less than 0.01% of the 685 Mb genome. Calculating the average SNP density across the genome:

$$(\text{Genome length in basepairs} / \text{Total number of shared SNPs})$$

suggested one SNP every 5.7 kb - assuming an even spacing of SNPs. The mean number of genotyped sites per RADlocus was 233.23 bp (standard error 0.14 bp). On average, 26.2/30 and 27.7/32 individuals were analysed per RADlocus, for each of the pre- and post-IVM treatment sample groups respectively. The nucleotide diversity ( $P_i$ ) for variant sites was very similar for each sample group, being 0.223 and 0.224 for the pre- and post-IVM populations respectively. The genome-wide nucleotide diversity was considerably lower being 0.002 for both samples. Not all alleles were present in both populations. The pre-treatment sample had 1365 private alleles, and the post-IVM sample contained 1768 private alleles. The mean  $F_{ST}$  between the pre- and post-treatment samples was 0.013.

#### **4.3.5 A diverse population but with the presence of contamination: kinship analysis**

High individual diversity was apparent following genome wide population analysis using the R program Adegnet v2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011). A genotype plot, displaying the SNP genotypes for all individuals sequenced was produced. By chance, due to random assortment of individuals by the software, this plot highlighted the presence of two highly similar individuals (Figure 4.5). Despite a high percentage of identical genotypes, the presence of a few notable differences between the two individuals was still apparent. One individual had been sampled pre-treatment and the other post-IVM. When the plot was re-constructed, with individuals sorted by treatment sample, no obvious segregation, suggestive of population sub-structuring or selection by IVM of SNP genotypes, was apparent (Figure 4.6).

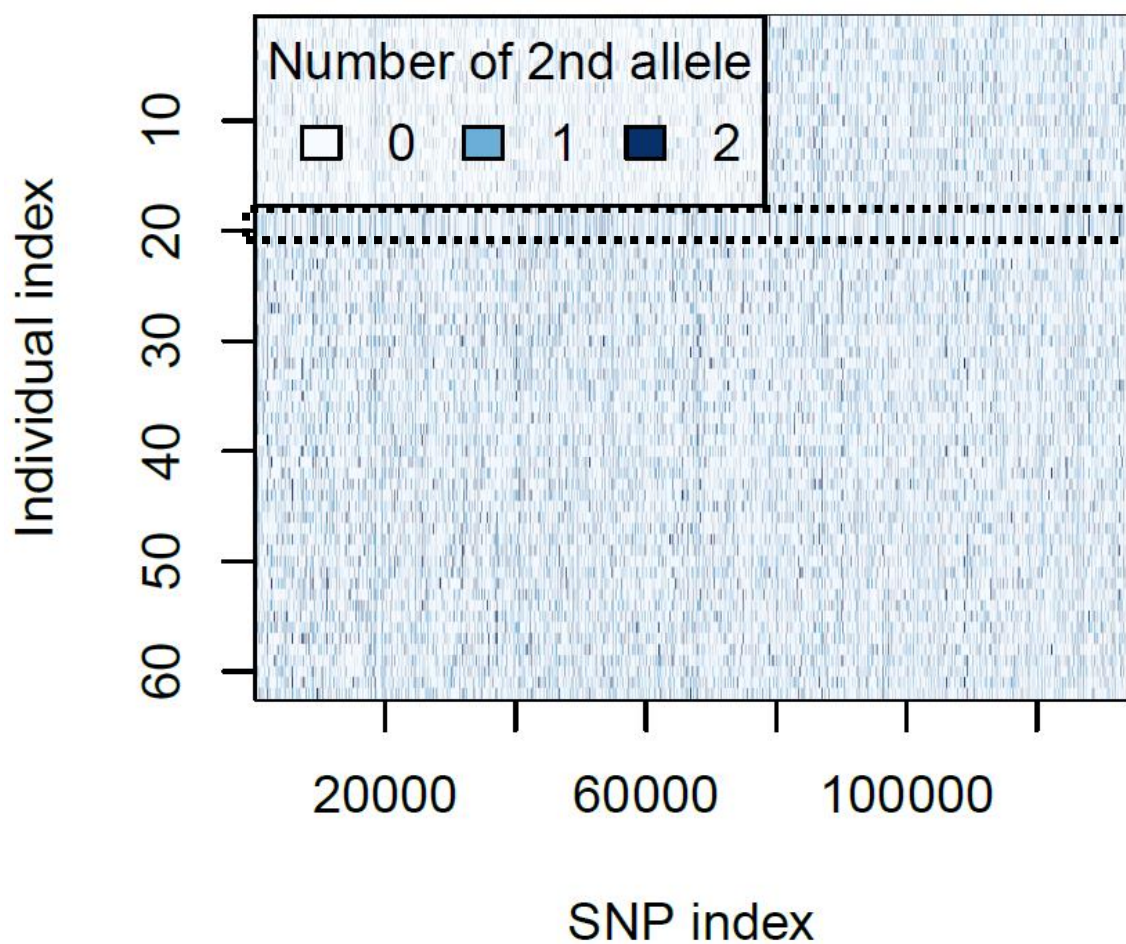
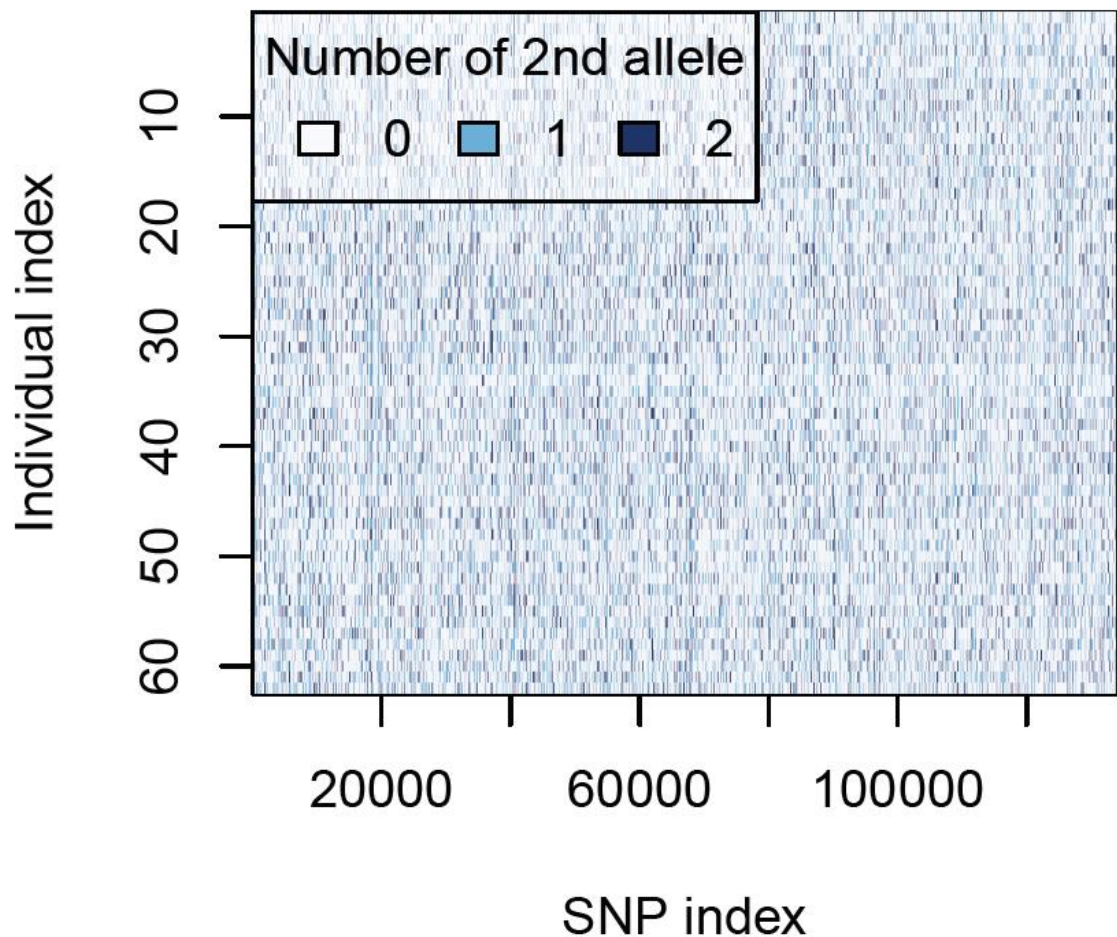
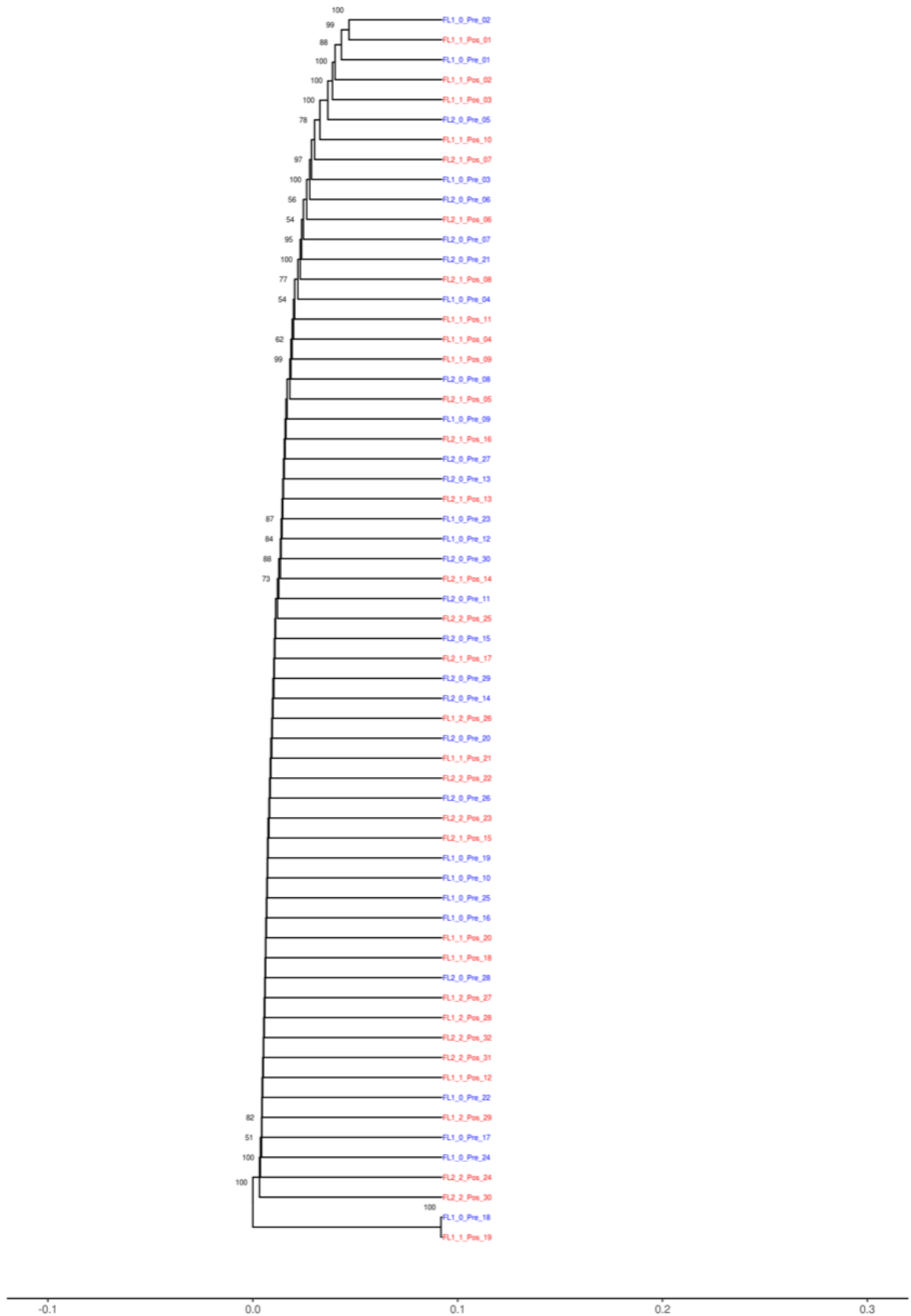


Figure 4.5: Genotype plot of all individuals. Each row is a separate individual. Outlined by a dotted box are two highly similar individuals. All SNPs identified are included, with each column corresponding to a different SNP position. The individual's genotype is indicated for each SNP by three colours; white = homozygous for reference allele (0), light blue = heterozygous (1), dark blue = homozygous for alternative allele (2).

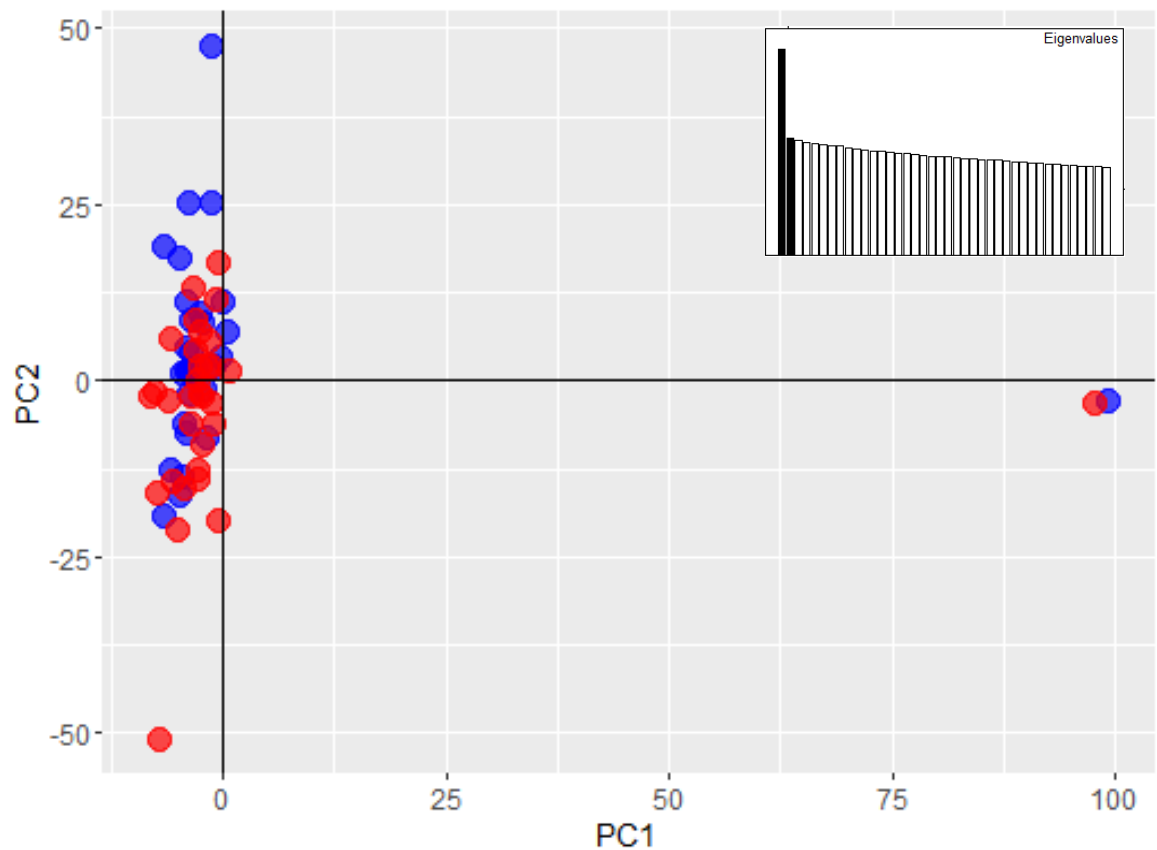


**Figure 4.6: Genotype plot of all individuals ordered by treatment sample. This is a different visualisation of the same information plotted in Figure 4.5. Each row is a separate individual. All SNPs identified are included, with each column corresponding to a different SNP position. The individual's genotype is indicated for each SNP by three colours; white = homozygous for reference allele (0), light blue = heterozygous (1), dark blue = homozygous for alternative allele (2).**

Phylogenetic analysis confirmed high individual diversity, for all but the two previously mentioned individuals (Figure 4.7). A PCA plot revealed that 3.69% of the overall variation could be explained by the presence of these two individuals, separated along the first principal component axis from all other individuals. The remaining individuals were spread along the second principal component axis, which described just 2.09% of all variation (Figure 4.8).



**Figure 4.7: Genetic distance tree of all 62 ddRAD-Seq individuals. An average neighbour method was used to cluster individuals. Key to colours: Blue = Pre-treatment, Red = Post-ivermectin. Scale = fraction of sites which differ between individuals (bitwise.dist()).**



**Figure 4.8: PCA plot of all individuals. Adegenet v2.1.1 was used to calculate principal components, and *ggplot2* (Wickham, 2016) was used to plot individuals, colouring by treatment sample. The plot shows the variance explained by the first two principal components. The inset shows the 61 eigenvalues, each explaining 3.69% or less of the overall variance. Key to colour: Blue = Pre-treatment, Red = Post-ivermectin.**

A kinship analysis was performed using *KING-robust* (Manichaikul et al., 2010). If individuals were considered to be from the same population the kinship coefficient would be close to zero. Figure 4.9 shows the presence of a pair of closely related individuals, which have a kinship coefficient of 0.497 suggestive of monozygotic twin familial status. However, all other individual pairwise comparisons had negative kinship coefficients indicating unrelated individuals, with some population stratification, although no sub-groupings were obvious, even when sample population information was provided.

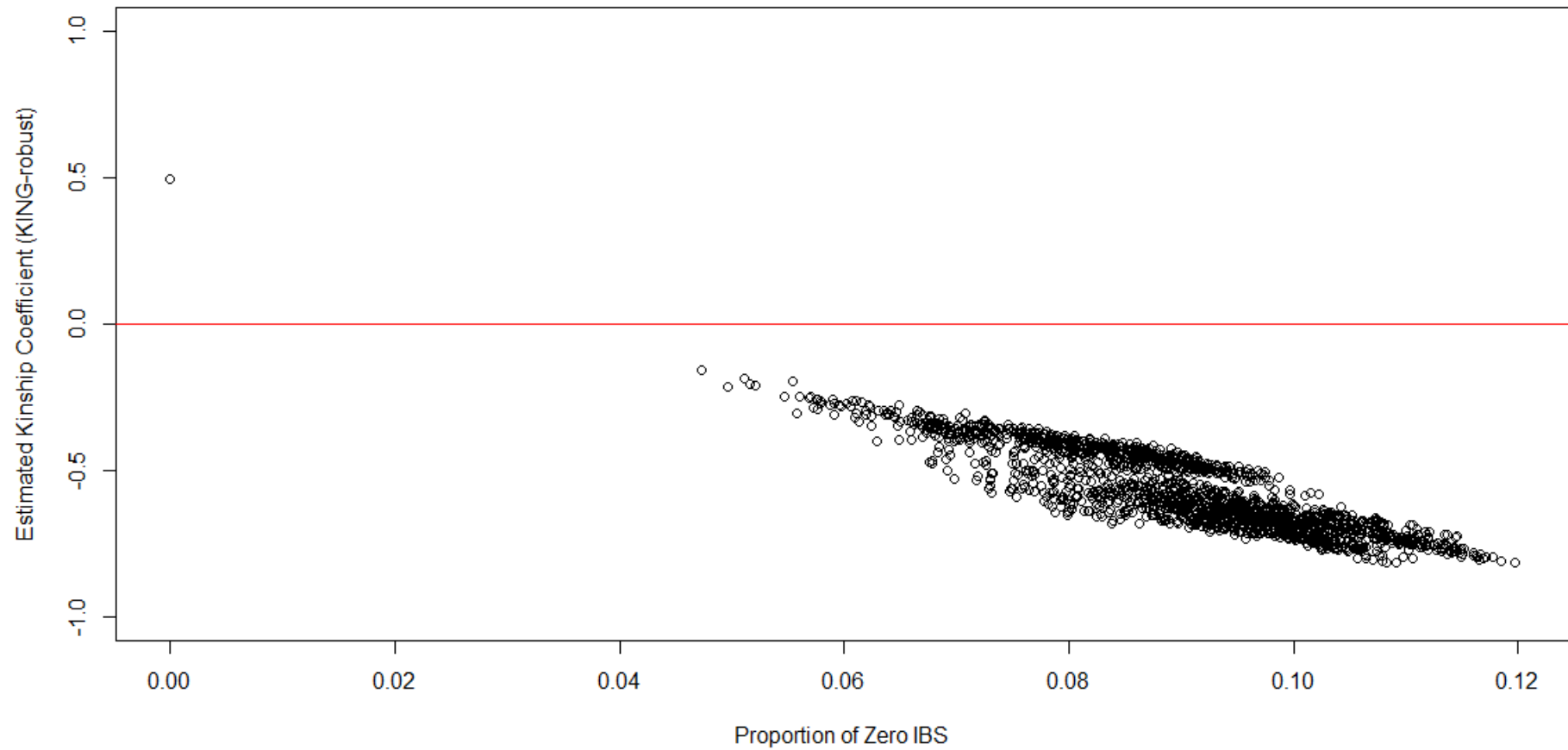
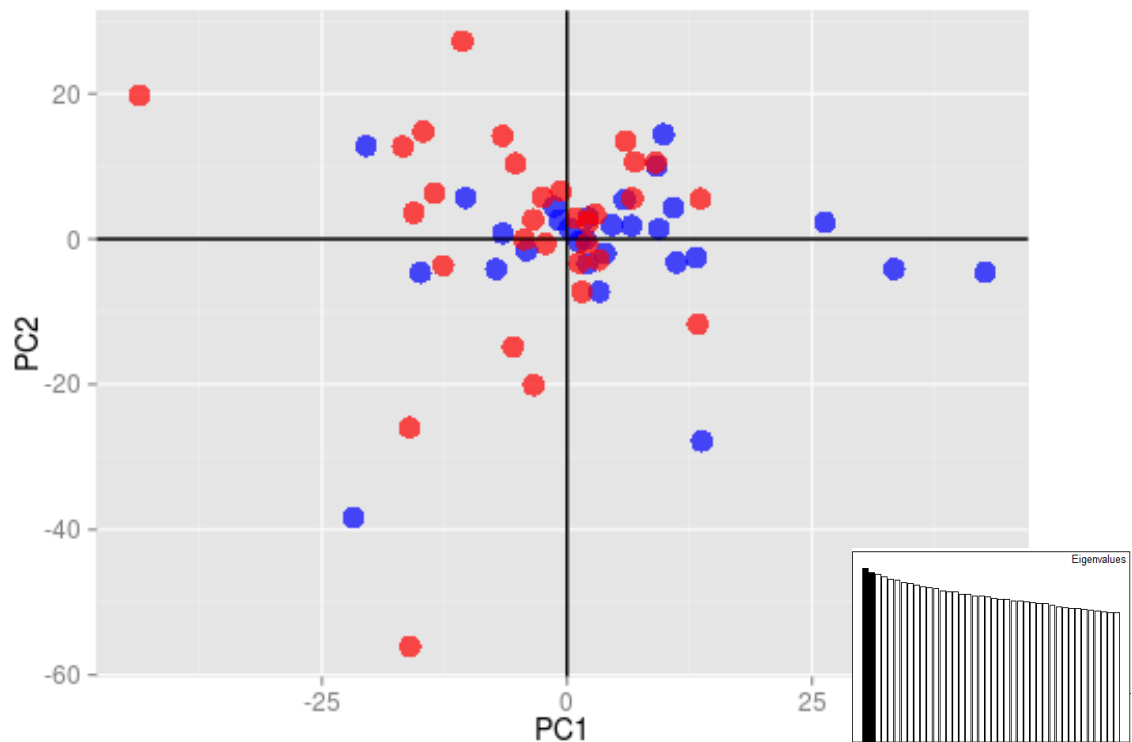


Figure 4.9: ddRAD-Seq kinship analysis of all 62 individuals. Using the *KING-robust* method in R (Manichaikul et al., 2010), kinship analysis between each all individuals was performed. Each point represents one comparison between a pair of individuals. If the kinship coefficient is estimated at zero, there is no relationship between the individuals, if at 0.5, then the individuals are monozygotic twins, and if negative then the individuals are unrelated, but there may be population sub-structuring. IBS = Identity by State.

The two highly similar individuals were identified as Pre\_18 and Post\_19. Following this analysis and examination of lab records, where it was shown that these two were prepared in neighbouring plate wells, it was decided to remove both individuals from further statistical analyses. Re-plotting the PCA now demonstrated the presence of a single population, with each of the 59 eigenvalues contributing 2.18% or less of the overall variation (Figure 4.10).



**Figure 4.10: PCA plot of 60 individuals. Following removal of the Pre-18 and Post-19 highly similar individuals, the PCA was repeated, and the plot shows the variance explained by the first two principal components. The inset shows the number of eigenvalues overall, each explaining 2.18% or less of the overall variance. Key to colour: Blue = Pre-treatment, Red = Post-ivermectin.**

#### 4.3.6 Many sites were out of Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium (HWE) describes the presence of entirely random mating within a population and requires genotype frequencies to match those expected given certain allele frequencies. If the expected genotypes (heterozygous/homozygous ratios) are not observed then it suggests that non-random mating has occurred, such as might be expected following selection and bottlenecking of the population as a result of IVM treatment. RADloci haplotypes were assessed for HWE using Stacks v2.0b (Rochette et al., 2019). It was determined that 8927 of the 15,809 variants containing RADloci (56.5%) in the

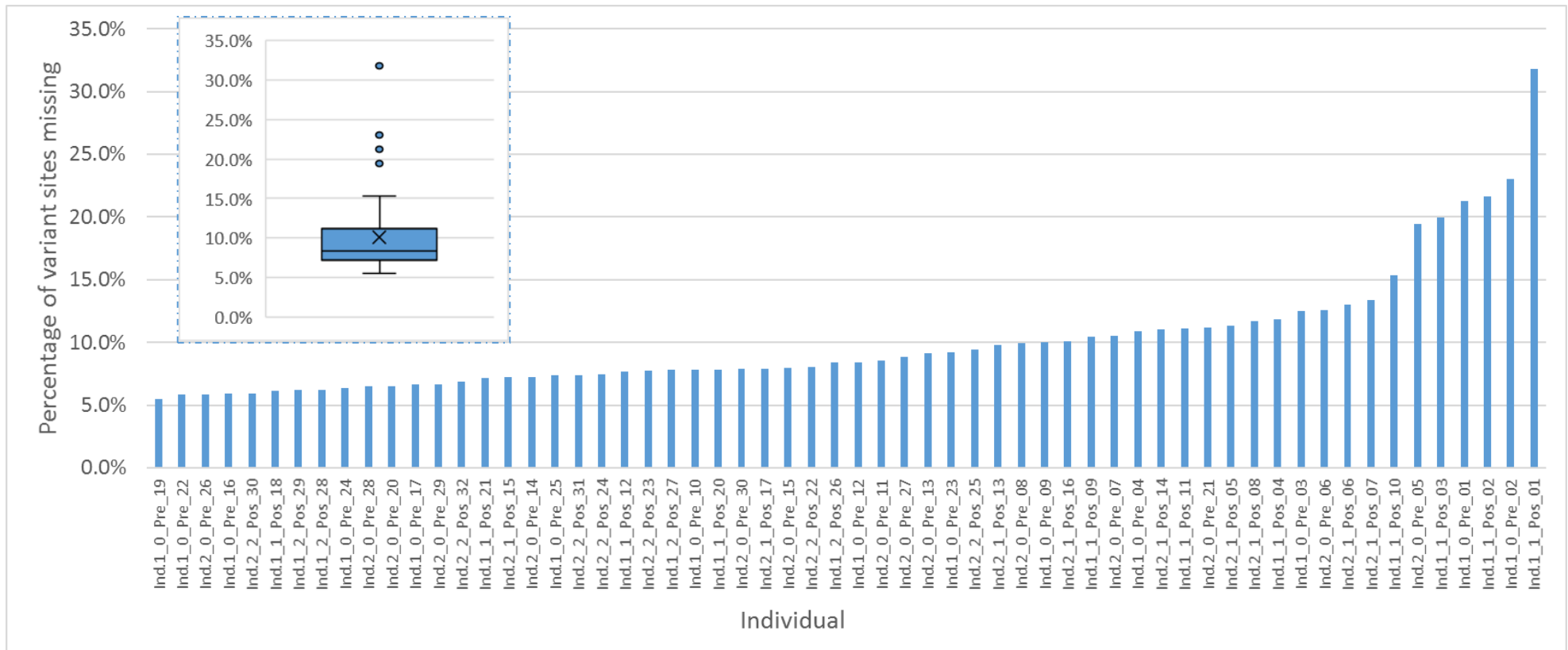


post-IVM sample group were significantly deviating from HWE ( $p < 0.05$ ). In the pre-treatment sample group fewer RADloci (15,773) contained variants, and 8622 (54.7%) of these RADloci deviated significantly from HWE ( $p < 0.05$ ).

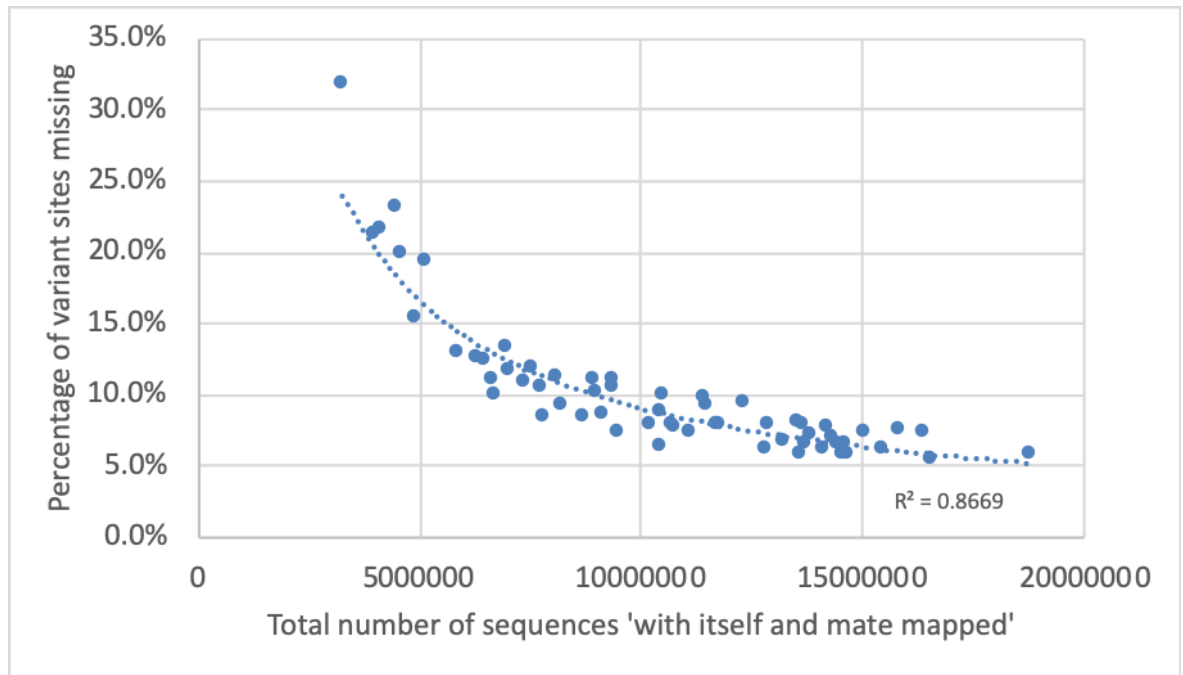
Both Stacks v2.0b and VCFtools v0.1.14 (Danecek et al., 2011) were used to analyse SNPs. Of the 119,892 variant sites, 49.5% (59,407) were deviating from HWE. By sample population, 27.2% (32,622) of pre-treatment and 29.2% (35,023) of post-IVM variant sites were deviating from HWE. Overall, observed heterozygosity of variant sites was lower than expected (0.15 observed heterozygosity for both pre- and post-IVM, as opposed to 0.22 expected heterozygosity for both). Almost all variant sites deviating from HWE, (59,323), were considered to be in heterozygous deficit ( $p < 0.05$ ). Just 84 variant sites (0.07% of all sites), were considered to be in heterozygous excess ( $p < 0.05$ ).

#### **4.3.7 Six individuals were missing proportionally more common RADloci than others**

Using VCFtools v0.1.14, the percentage of missing variant sites per individual was calculated, compared to the total number of variant sites retained across all 60 individuals (Figure 4.11). The individual with the lowest percentage of sites missing had 5.5% of variant sites missing, while one individual had as many as 31.8% of variant sites missing from their data. On average each individual had 10.1% missing data, however the median was lower (8.4%) due to skewing of the data by six individuals, shown as outliers in the inset in Figure 4.11. Comparing this data with the number of sequences aligned as a pair to the genome, it was clear that there was a negative correlation between the percentage missingness and the total number of sequences aligned (Figure 4.12).



**Figure 4.11: Percentage of shared variant sites missing per individual. Six individuals were outliers (see inset), with a higher percentage of sites missing, such that the mean (X), is greater than the median.**



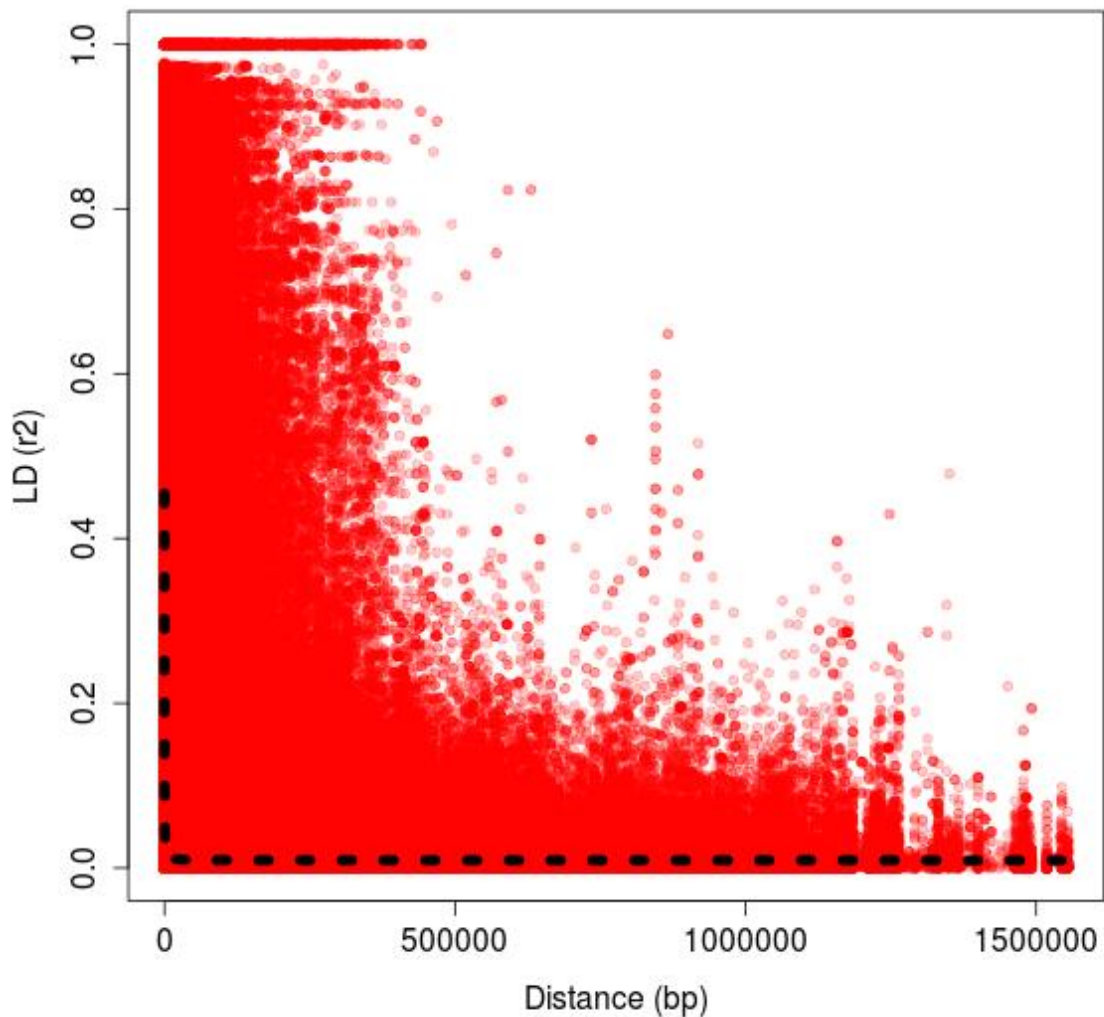
**Figure 4.12: Percentage of variant sites missing per individual compared with sequences aligned. Each dot represents one individual. Using Excel, a regression line (power) was fitted,  $R^2=0.8669$ . The sequences included are those which are mapped as a pair to the genome, although not necessarily mapped together in their pair ('properly paired') in mapping.**

#### 4.3.8 Linkage disequilibrium decay was rapid

Following the removal of the six individuals with the highest percentage of missing sites, Stacks v2.0b *populations* was repeated and VCFtools v0.1.14 was used to calculate the LD. The squared correlation coefficient value ( $r^2$ ) between each pair of variant genotypes on a contig was chosen. The squared correlation coefficient,  $r^2$ , is allele frequency dependent, and has a range of zero to one. It will only be high if both alleles are at the same allele frequency and linked. Otherwise, if one allele is rare, even though it is linked with the other, the value of  $r^2$  will always be low. This makes it a more useful measure of LD within the genome for this study than  $D'$ , the alternative measure of LD, as this latter estimate will be high if alleles are linked, regardless of their frequency within the population. As such,  $r^2$  will only suggest high LD if alleles are both common and linked (which is required for identification of genetic markers linked to a causal mutation), with the potential for use in other populations.

Distances between genotype positions ranged from 1 bp to 1.55 Mbp, and the mean  $r^2$  value was 0.090 (range 0.000 - 1.000), while the median value was much lower - only 0.016. The half-decay distance was always the minimum distance

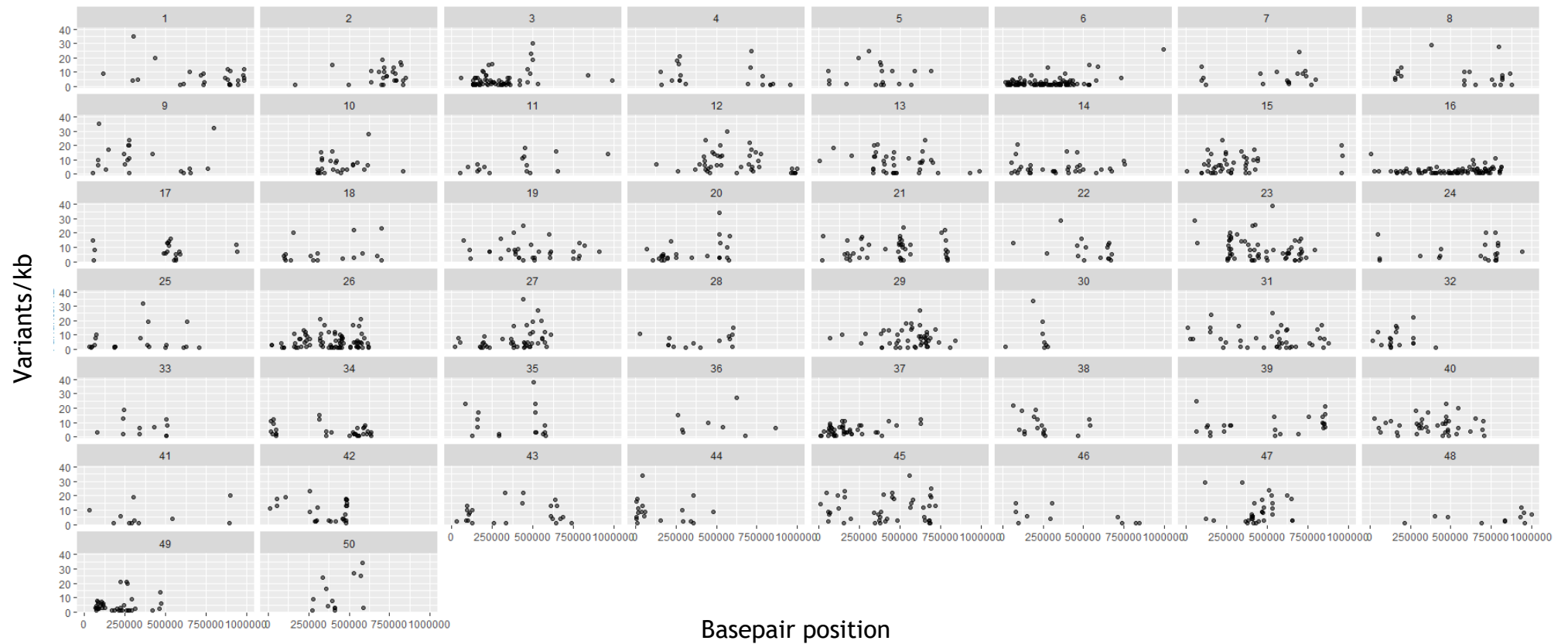
between positions, and so, as a more meaningful measure, the distance to the mean intercontig LD (0.0206) was calculated (Cutter et al., 2006), and was found to be 2805 bp. Although some genotypes appeared to be linked over a considerable distance, the majority had very low  $r^2$  values, even over short distances, and the linkage decay was extremely rapid (Figure 4.13). This data will necessarily be affected by the lengths of contigs within the Tci2 genome. The average contig length is only 85.5 kb, but many contigs are only a few thousand basepairs long. If LD was high on short contigs, but lower on longer contigs, this could overestimate the speed of linkage decay and underestimate the half decay distance.



**Figure 4.13: Linkage decay plot.** Each point represents the LD ( $r^2$ ) between one pair of variant genotypes, within the same contig. The distance in basepairs between genotype positions is plotted along the x-axis. The linkage decay is plotted as a black dotted line and falls sharply.

#### 4.3.9 Low and variable SNP density across the genome

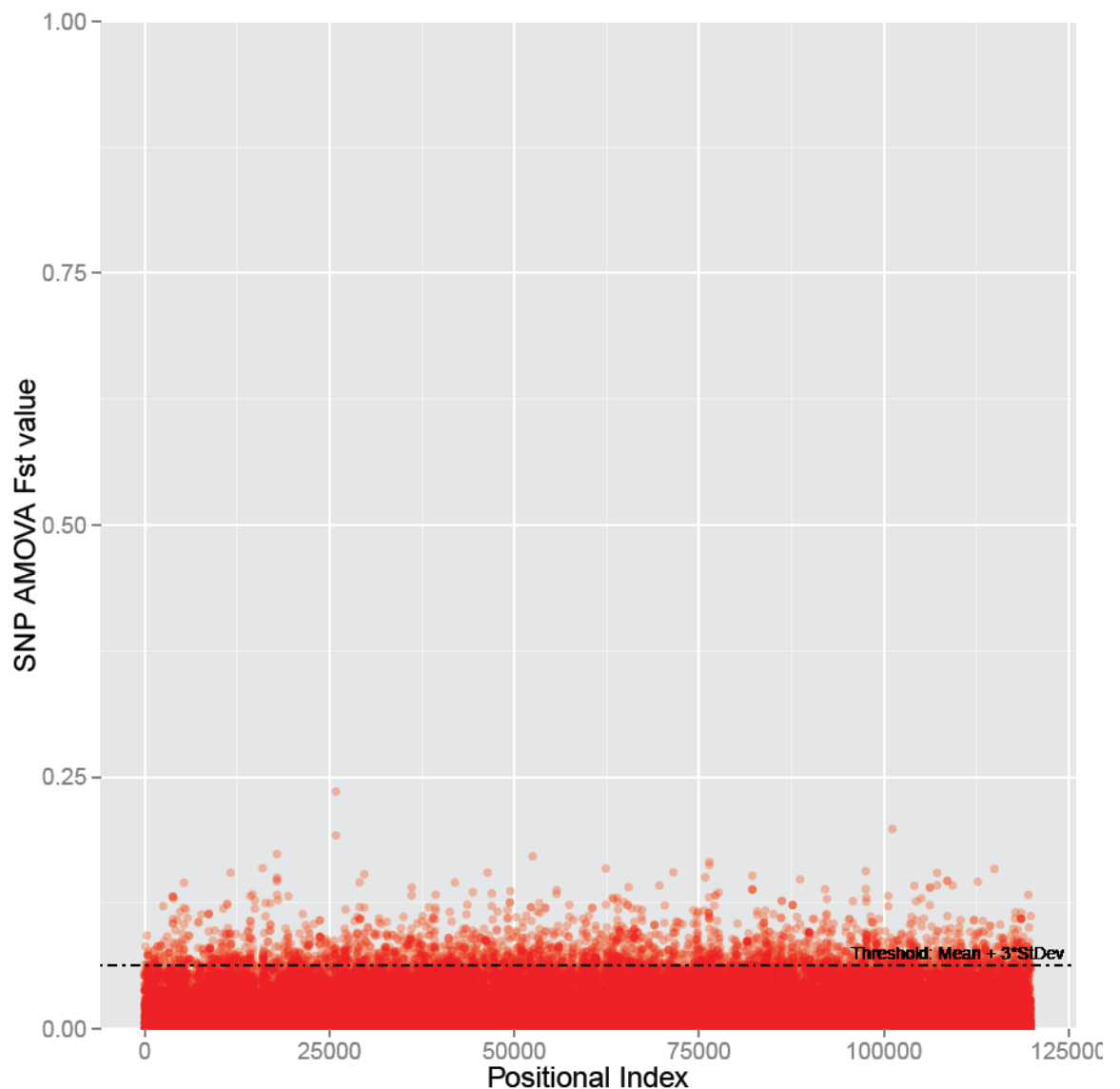
VCFtools v0.1.14 was used to assess the SNP density per kb across the genome and contigs 1 to 50 in the Tci2 genome were plotted (Figure 4.14). These were generally the longest contigs, with the first contig measuring 2.08 Mb, however, not all contigs in the Tci2 genome are in size order. From these plots it was clear that the SNP density per kb varied widely across the genome, including within a contig, and there were large regions where no RAD markers were present (Figure 4.14).



**Figure 4.14: ddRAD-Seq SNP density across the genome. The first 50 contigs are plotted, each on a separate plot. Each dot represents a 1 kb window, which contains at least one variant site. For each contig the first million bp are included, however some contigs shown will be shorter in length than this.**

### 4.3.10 $F_{ST}$ analysis identifies genetically differentiated RADloci as potential markers of IVM resistance

A Manhattan plot of the  $F_{ST}$  values for each SNP identified a subset of highly differentiated SNPs but did not reveal any obvious clustering (Figure 4.15). This is likely to reflect the fragmented state of the genome (Doyle et al., 2019). Filtering the dataset using a threshold ( $F_{ST} = 0.0637$ ) of three standard deviations above the mean  $F_{ST}$  value retained 2592 RAD markers (Appendix 4).



**Figure 4.15: ddRAD-Seq SNP  $F_{ST}$  Manhattan plot. All SNPs are plotted by positional index within the file rather than by their genomic location. However, the SNPs are in their positional order from the start of contig 1 (left) to the end of the last contig sampled (right). The dotted line indicates the threshold of three standard deviations above the mean  $F_{ST}$ . (Threshold = 0.0637).**

$F_{ST}$  values indicate the degree of fixation of an allele within a population, relative to another population, but any within-sample diversity greatly reduces  $F_{ST}$  (Bird et al., 2011). Standardization of an  $F_{ST}$  value ( $F'_{ST}$ ) to account for within-sample diversity can help to recognise allele selection in populations with high individual diversity (i.e. given the diversity present, what is the maximum possible  $F_{ST}$  value attainable and, of this, what is the relative  $F_{ST}$  value?) (Bird et al., 2011). A true measure of between-population differentiation is  $D_{EST}$ , which is independent of within-population diversity (Bird et al., 2011). It was decided to use all three measures to identify potential loci under IVM selection. It should be noted that while the  $F_{ST}$  value is calculated per SNP, the  $F'_{ST}$  and  $D_{EST}$  values are calculated per RADlocus and are thus comparing RADlocus haplotypes, effectively acting as a local smoothing measure for SNP analysis. Although the smoothed AMOVA  $F_{ST}$  value (using 1 kb distinct windows) was also extracted, the regional spread of SNPs across the genome varied greatly and had a noticeable impact on the values calculated. Since the robustness of this algorithm was questionable in this particular analysis, it was not considered as important as the  $F_{ST}$ ,  $F'_{ST}$  and  $D_{EST}$  values.

As the genome was not annotated at the time of analysis, it was decided to manually annotate regions of the genome surrounding SNPs of interest. Annotating all contigs containing SNPs greater than the threshold was considered unmanageable, but there were no obvious clusters of SNPs greater than the threshold, which could be visually located. Therefore, RADloci were chosen for further investigation as described in 2.7.8. In total, 19 SNPs in 17 RADloci, each on a separate Tci2 contig, were investigated further. Eleven RADloci were located within genes (Table 4.2). Eight RADloci were within 10 kb of at least one gene (Table 4.3). A further 27 putative genes were annotated within 50 kb of a RADlocus (Table 4.4), and finally an additional seven genes were annotated within 75 kb of a RADlocus of interest (Table 4.5). No candidate genes previously associated with IVM resistance were identified.



**Table 4.2: Genes containing a RADlocus of interest identified following manual annotation of selected high  $F_{ST}$  contigs.**

RADlocus	Tci2 contig	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 contig	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	<i>Caenorhabditis elegans</i> Chromosome	$F_{ST}$ value <sup>1</sup>
349441	77	TELCIR_09574	2025	HCON_00189500	R12E2.2 ( <i>suco-1</i> )	I	0.159
587481	146	TELCIR_11794	3246	HCON_00151030	M04G12.1 ( <i>tag-260</i> )	V	0.236
1486836	545	TELCIR_04325	421	HCON_00074220	K01A2.11 ( <i>cbn-1</i> )	II	0.159
1572681	598	TELCIR_06418	900	HCON_00041530	NA	NA	0.141
1759024	732	TELCIR_02631	169	HCON_00134530	R13H9.5	IV	0.155
1967245	894	NA	4362	HCON_00102640	NA	NA	0.089 <sup>2,3</sup>
2034639	953	TELCIR_04965	544	HCON_00167480	R01E6.6 ( <i>glb-20</i> )	X	0.152
2284168	1226	TELCIR_00369	11	HCON_00160120	F08F3.10	V	0.097 <sup>2,3</sup>
2464065	1471	TELCIR_11681 TELCIR_11682	3169	HCON_00013050	W01A8.5 ( <i>tofu-5</i> )	I	0.157
2581282	1680	TELCIR_01924	103	HCON_00079090	K01B6.2 ( <i>srx-45</i> )	III	0.198
2976600	2896	TELCIR_13774 TELCIR_13775	4779	HCON_00034960	Y38H6C.17 H32K16.1 ( <i>slc-36.3</i> , <i>slc-36.4</i> )	V, I	0.146

These RADloci were within the top 50  $F_{ST}$  results or were within the top 50  $F'_{ST}/D_{EST}$  results. Genes were manually annotated in Artemis using RNA-Seq data (Choi et al., 2017), and equivalent proteins and homologues identified by BLASTP of the amino acid sequence. PRJNA72569 contigs were identified using TBLASTN when no equivalent protein was identified. <sup>1</sup>The  $F_{ST}$  value of the highest ranking SNP with the RADlocus is provided. <sup>2</sup>Selected based on  $F'_{ST}$  results. <sup>3</sup>Selected based on  $D_{EST}$  results.

**Table 4.3: Genes within 10 kb of a RADlocus of interest identified following manual annotation of selected high  $F_{ST}$  contigs.**

RADlocus	Tci2 contig	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 contig	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	<i>Caenorhabditis elegans</i> Chromosome	$F_{ST}$ value <sup>1</sup>
243878	50	NA	8051	HCON_00104180	C28C12.7 ( <i>spp-10</i> )	IV	0.100
243878	50	TELCIR_13282	4361	HCON_00104170	T05G5.3 ( <i>cdk-1</i> )	III	0.100
384468	86	TELCIR_04598 TELCIR_04599	477	HCON_00073720	F25B5.1 ( <i>dct-6</i> )	III	0.064 <sup>3</sup>
513975	122	TELCIR_00464	14	HCON_00160450	ZC116.3	V	<0.0637 <sup>2,3</sup>
643960	163	NA	4878, 132	HCON_00148790	C24G6.3 ( <i>mms-19</i> )	V	0.145
643960	163	TELCIR_13875 TELCIR_02259 TELCIR_02260	4878, 132	HCON_00148780	C24G6.6 ( <i>hpo-15</i> )	V	0.145
1170167	373	TELCIR_01410	357, 2244, 62	NA	NA	NA	0.137
1237196	404	TELCIR_05309	622	HCON_00189240	F39B1.1 ( <i>piki-1</i> )	X	0.171
1486836	545	NA	343494	NA	NA	NA	0.159
1486836	545	TELCIR_04326	421	HCON_00074210	H14N18.3 ( <i>ttr-47</i> )	V	0.159
2464065	1471	TELCIR_11682	3169	HCON_00013040	Y60A3A.12 ( <i>chk-2</i> )	V	0.157

These RADloci were within the top 50  $F_{ST}$  results or were within the top 50  $F'_{ST}/D_{EST}$  results. Genes were manually annotated in Artemis using RNA-Seq data (Choi et al., 2017), and equivalent proteins and homologues identified by BLASTP of the amino acid sequence. PRJNA72569 contigs were identified using TBLASTN when no equivalent protein was identified. <sup>1</sup>The  $F_{ST}$  value of the highest ranking SNP with the RADlocus is provided. <sup>2</sup>Selected based on  $F'_{ST}$  results. <sup>3</sup>Selected based on  $D_{EST}$  results.

Table 4.4: Genes within 50 kb of a RADlocus of interest identified following manual annotation of selected high  $F_{ST}$  contigs.

RADlocus	Tci2 contig	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 contig	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	<i>Caenorhabditis elegans</i> Chromosome	$F_{ST}$ value <sup>1</sup>
243878	50	TELCIR_09426 TELCIR_09425	1952	HCON_00104140 HCON_00104160	Y116A8C.36 ( <i>itsn-1</i> )	IV	0.100
384468	86	TELCIR_26237 TELCIR_23945	375630, 205630	HCON_00073750	H26D21.1 ( <i>hus-1</i> )	I	0.064 <sup>3</sup>
384468	86	TELCIR_26347	494832	HCON_00073740	T12D8.6 ( <i>mlc-5</i> )	III	0.064 <sup>3</sup>
389043	86	TELCIR_09534	2004	HCON_00148060	Y32F6A.4	V	0.174
389043	86	TELCIR_19868	15459, (3079, no genes)	HCON_00148050	NA	NA	0.174
587481	146	TELCIR_15805	6898	HCON_00151040	T01D3.5 ( <i>zipt-9</i> )	V	0.236
587481	146	TELCIR_10786	2599	HCON_00151060	R186.3	V	0.236
643960	163	TELCIR_02258	132	HCON_00148760	C24G6.4 ( <i>nhr-47</i> )	V	0.145
1486836	545	TELCIR_01295 TELCIR_04328	54, 421	NA	Y54F10AM.2 ( <i>feh-1</i> )	III	0.159
1572681	598	TELCIR_06417	900	HCON_00041530	NA	NA	0.141
1759024	732	TELCIR_02631	169	HCON_00134530	C39H7.1	IV	0.155
1759024	732	NA	169	HCON_00134530	C39H7.1	IV	0.155

RADlocus	Tci2 contig	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 contig	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	<i>Caenorhabditis elegans</i> Chromosome	F <sub>ST</sub> value <sup>1</sup>
1886869	827	NA	426	NA	NA	NA	0.166
1886869	827	NA	NA	NA	NA	NA	0.166
1886869	827	TELCIR_05657	699	HCON_00178015	R148.3	III	0.166
1886869	827	TELCIR_05658 TELCIR_05659 TELCIR_05660	699	HCON_00179670	F48E3.3 F26H9.8 ( <i>uggt-1</i> , <i>uggt-2</i> )	X, I	0.166
1967245	894	TELCIR_06338	870	NA	NA	NA	0.089 <sup>2,3</sup>
1967245	894	NA	3253, 870	HCON_00096840	Y41E3.9 ( <i>fcd-2</i> )	IV	0.089 <sup>2,3</sup>
2034639	953	NA	544	NA	C56E6.4	II	0.152
2284168	1226	TELCIR_00368	11	HCON_00160130	C01G10.8 ( <i>ahsa-1</i> )	V	0.097
2464065	1471	TELCIR_03475	280	HCON_00013030	M01E11.2	I	0.157
2581282	1680	TELCIR_17075 TELCIR_02329	138	HCON_00091250	K08E3.5	III	0.198
2748445	2050	TELCIR_12625 TELCIR_12624	3834	HCON_00066190	Y50D7A.4 ( <i>hpo-29</i> )	III	<0.0637 <sup>2</sup>
2976600	2896	TELCIR_18588	11602	HCON_00153600	F47G9.4	V	0.146
3071877	3471	TELCIR_07163	1119	HCON_00071130	Y53G8AR.8	III	0.159
3071877	3471	TELCIR_07164	1119	HCON_00071130	ZK688.8 ( <i>gly-3</i> )	III	0.159

RADlocus	Tci2 contig	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 contig	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	<i>Caenorhabditis elegans</i> Chromosome	F <sub>ST</sub> value <sup>1</sup>
3071877	3471	TELCIR_07165 TELCIR_14502	5449, 1119	HCON_00071140	C29E4.5 ( <i>tag-250</i> )	III	0.159

These RADloci were within the top 50 F<sub>ST</sub> results or were within the top 50 F'<sub>ST</sub>/D<sub>EST</sub> results. Genes were manually annotated in Artemis using RNA-Seq data (Choi et al., 2017), and equivalent proteins and homologues identified by BLASTP of the amino acid sequence. PRJNA72569 contigs were identified using TBLASTN when no equivalent protein was identified. <sup>1</sup>The F<sub>ST</sub> value of the highest ranking SNP with the RADlocus is provided. <sup>2</sup>Selected based on F'<sub>ST</sub> results. <sup>3</sup>Selected based on D<sub>EST</sub> results.

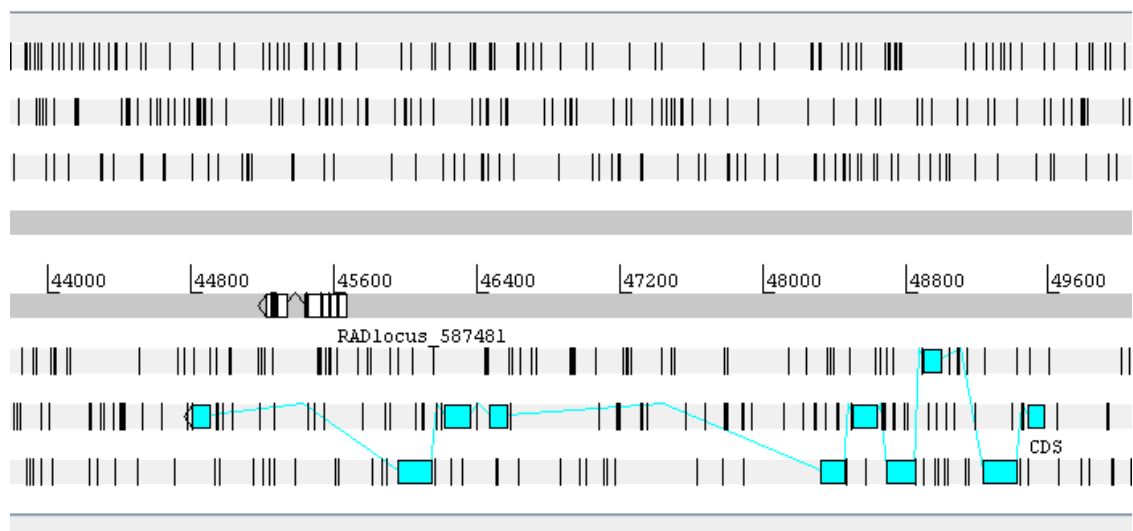
**Table 4.5: Genes within 75 kb of a RADlocus of interest identified following manual annotation of selected high  $F_{ST}$  contigs.**

RADlocus	Tci2 contig	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 contig	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	<i>Caenorhabditis elegans</i> Chromosome	$F_{ST}$ value <sup>1</sup>
349441	77	TELCIR_06342	872	HCON_00189530	B0395.1 ( <i>nhx-1</i> )	X	0.159
389043	86	TELCIR_05913	759	HCON_00148110	NA	NA	0.174
643960	163	TELCIR_08013	1395	HCON_00148730	T05C3.5 ( <i>dnj-19</i> )	V	0.145
1170167	373	TELCIR_08843 TELCIR_18188	1715, 10765	HCON_00106310	T06E4.1	V	0.137
1170167	373	NA	NA	NA	NA	NA	0.137
1170167	373	TELCIR_19422	13874	HCON_00101380	M03F4.2 T04C12.4 T04C12.6 ( <i>act-1</i> , <i>act-3</i> , <i>act-4</i> )	X, V, V	0.137
2034639	953	TELCIR_03238 TELCIR_06860 TELCIR_08734	177, 1025, 789, 244, 1672	HCON_00046190	ZK20.6 ( <i>nep-1</i> )	II	0.152

These RADloci were within the top 50  $F_{ST}$  results or were within the top 50  $F'_{ST}/D_{EST}$  results. Genes were manually annotated in Artemis using RNA-Seq data (Choi et al., 2017), and equivalent proteins and homologues identified by BLASTP of the amino acid sequence. <sup>1</sup>The  $F_{ST}$  value of the highest ranking SNP with the RADlocus is provided.

### 4.3.11 RADlocus 587481 had the highest $F_{ST}$ value SNP, yet on closer inspection did not appear to be under IVM selection.

The SNP with the highest  $F_{ST}$  value (0.236) was located within RADlocus 587481 on Tci2 contig 146. Close to this SNP, within the same RADlocus, was a second SNP, also estimated to have a high  $F_{ST}$  value. In addition, this RADlocus was within the top 14 RADloci assessed by both  $F'_{ST}$  and  $D_{EST}$ . Despite this, these two SNPs were not in the top 50 smoothed AMOVA  $F_{ST}$  results. The RADlocus was found to reside within the 9<sup>th</sup> intron of a ten exon annotated gene (Table 4.2), with genomic location Tci2 contig 146:44804-49573, the contig being 1.22 Mb in length (Figure 4.16).



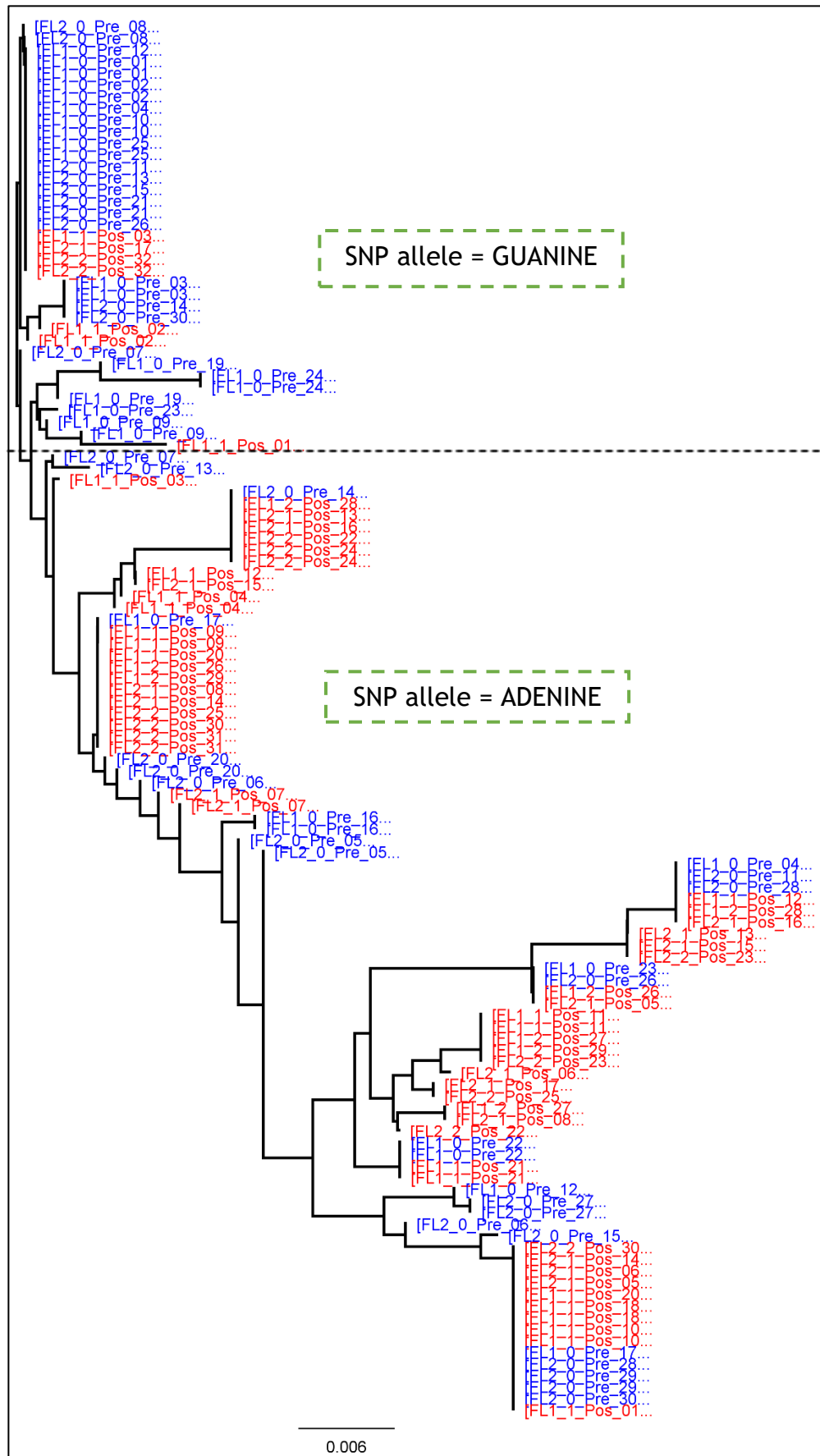
**Figure 4.16: RADlocus 587481 position.** Using RNA-Seq data (Choi et al., 2017), genes were annotated in Artemis and the RADlocus was found to reside within the 9<sup>th</sup> intron of a gene on contig 146 as shown. The exons are coloured blue, the RADlocus is white.

The highest  $F_{ST}$  SNP had either an adenine or a guanine encoded, and these alleles were in roughly equal proportions in the pre-treatment sample (47 and 53%). In the post-IVM sample just 9% of alleles were guanine, with the remaining 91% encoding adenine, suggesting positive selection of this allele.

Following extraction of the consensus sequenced RADtags for each individual for this locus, these were aligned in Geneious v9.1.8 (<https://www.geneious.com>), and the two SNPs with the high  $F_{ST}$  values were located. It became clear that these SNPs separated a common haplotype (containing guanine at position 50 and thymine at position 94) from other, less common, haplotypes within the

sample groups (Figure 4.17). Nevertheless, most individuals encoding this common haplotype were in the pre-treatment sample, with a diverse range of haplotypes present in the post-IVM sample. These multiple diverse haplotypes, present in both sample groups, encoded adenine at position 50 and cytosine at position 94 within the RADlocus. It therefore seemed unlikely that this RADlocus was under selection by IVM in this farm population.





**Figure 4.17:** A Neighbour-Joining tree of the consensus fasta sequences of RADlocus 587481, of all 60 individuals. Key to colours: blue = pre-treatment, red = post-IVM. Those in bold font, above the dashed line, encode guanine at the highest  $F_{ST}$  SNP. Those below the line encode adenine. Genetic distance model: Tamura-Nei, Outgroup: FL1\_1\_Pos\_01.

Plotting additional variants along Tci2 contig 146 showed that this RADlocus was unusual in displaying a higher level of SNP allele fixation between the pre- and post-treatment samples (Figure 4.18). Although there were other variant sites along the contig, none were as significant as these two SNPs, or had as high  $F_{ST}$  values. In addition, the majority of the other variant sites within the RADlocus, or close to it, had quite low  $F_{ST}$  values. Analysing the LD along the contig found it decayed as rapidly as the genome wide estimate of LD, and the distance in basepairs to the mean intercontig equilibrium was only 1958 bp.

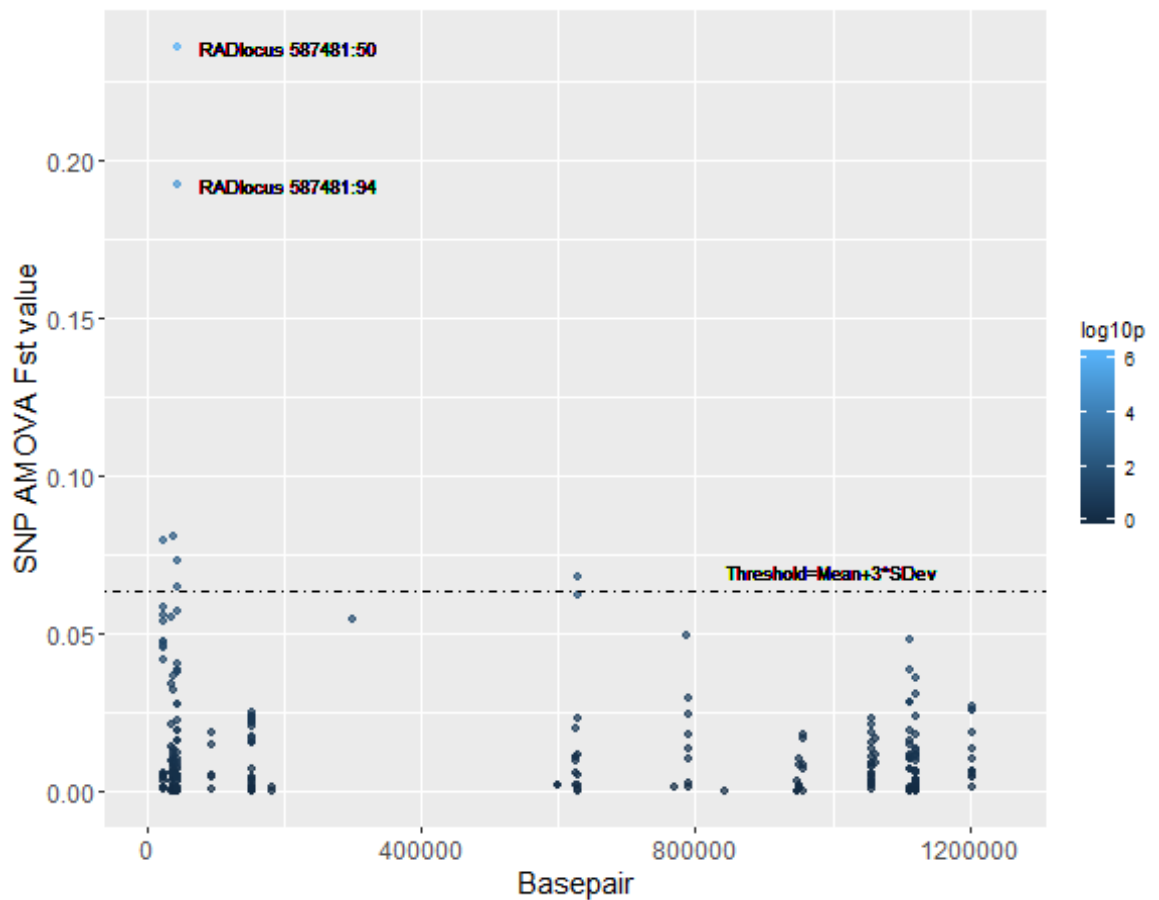
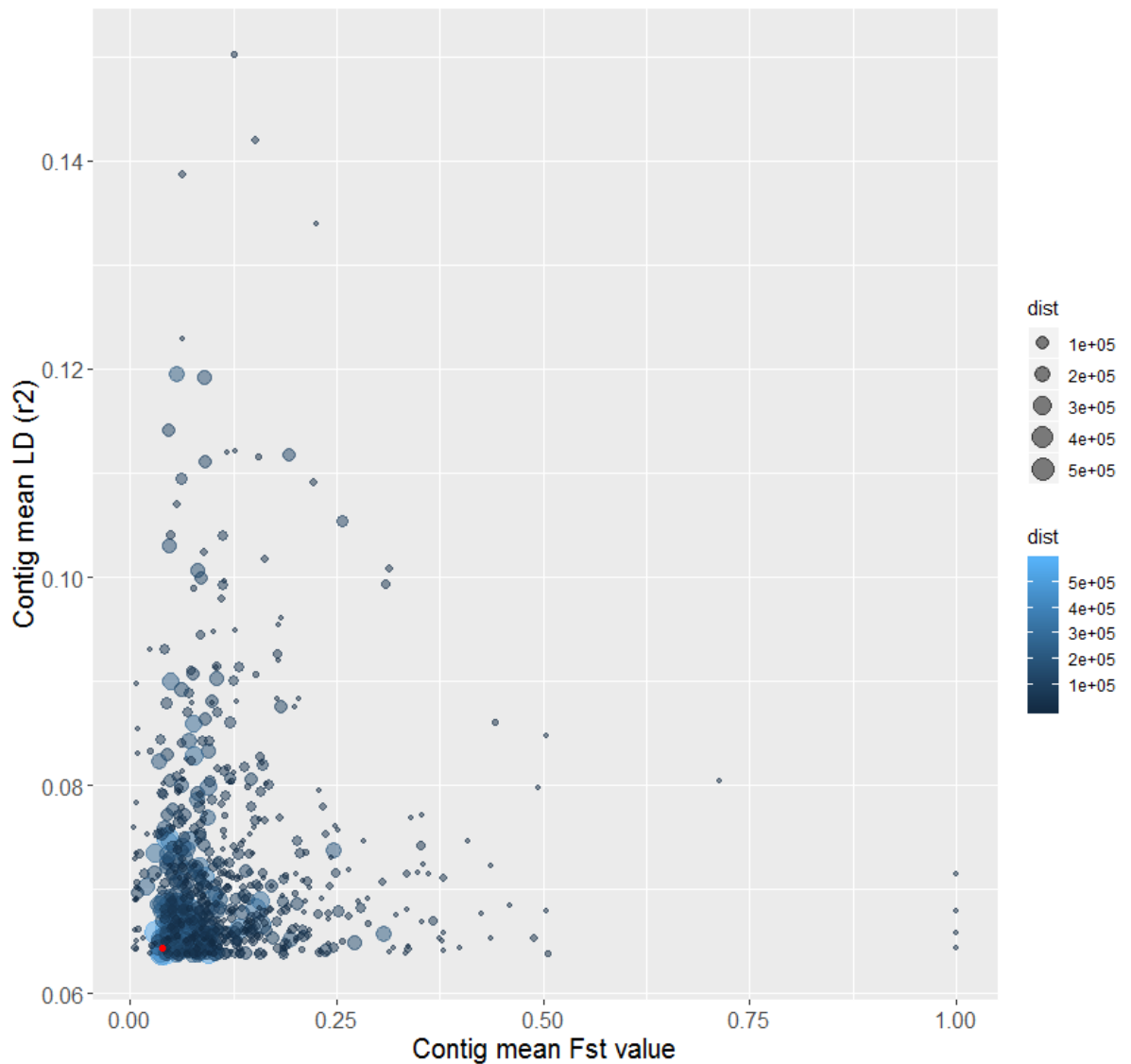


Figure 4.18: Contig 146 variant sites. All variant sites as provided by Stacks v2.0b were plotted by basepair position along contig 146. The  $F_{ST}$  threshold (three standard deviations greater than the mean) is indicated by the dot-dashed line. Each variant site represented is coloured by  $-(\log_{10}[p\text{-value}])$ . Lighter colours are more significant.

### 4.3.12 Linkage disequilibrium across high $F_{ST}$ contigs

Following analysis of RADlocus 587481 an alternative standard to judge whether or not to further investigate a genetically differentiated RADlocus was looked for. It was decided to assess whether any SNPs above the  $F_{ST}$  threshold were found on contigs of extensive LD. It was considered that such contigs might indicate selection by IVM and would be worth more intensive investigation. Data for contigs containing SNPs of  $F_{ST}$  value greater than the threshold were extracted. The LD decay was rapid for this subset of contigs, with the distance to the mean intercontig LD, 2831 bp, very similar to the genome-wide value. For these contigs, the mean  $r^2$  value was plotted against the mean  $F_{ST}$  value for each contig (Figure 4.19). Each was coloured and scaled such that the mean distance between SNPs on the contig was also visualised. From this it became clear that there was no correlation between contig mean  $F_{ST}$  value and contig mean  $r^2$  value, making it an unhelpful indicator to use for following up F-statistic analysis. Although this makes it unlikely that large regions are under selection by IVM, it does not rule out the presence of smaller regions within these contigs, especially within longer contigs, or misassembled contigs, when diversity around the RADlocus would reduce both contig mean  $F_{ST}$  and  $r^2$  values. Nevertheless, it was realised that contig 146 had both a comparatively low mean  $r^2$  and a low mean  $F_{ST}$  value.



**Figure 4.19: Contig mean  $F_{ST}$  value compared with mean  $r^2$  value.** Each of the contigs containing a RADlocus with a high  $F_{ST}$  value, greater than three standard deviations above the mean  $F_{ST}$  were extracted. The mean linkage disequilibrium  $r^2$  value was obtained for each contig, and the results are plotted here. For each contig the mean distance between base pairs used to calculate the linkage disequilibrium is represented both by the size of the plotted point, and by the colour, with larger distances being shown by larger, lighter coloured points. Contig 146 is shown in red.

## 4.4 Discussion

The aim of this experiment was to identify markers of IVM resistance in a field population of *T. circumcincta*. As we were working with a commercial farm, offspring of adults exposed to IVM were collected pre- and post-IVM treatment as part of a FECRT. A reduced representation method was chosen to allow an increased number of individuals to be sequenced within budget limitations. This generated individual genotypes with which we were able to assess population

structure, estimate LD and analyse samples for genetically differentiated variants, which could be used as markers of IVM resistance.

#### **4.4.1 Pre- and post-IVM sample populations appear as a single population**

The ddRAD-Seq confirmed the presence of a largely homogeneous population present on a UK commercial farm, as was expected from previous microsatellite analysis of *T. circumcincta* UK populations (Grillo et al., 2007; Redman et al., 2015). Kinship analysis allowed identification of potential contamination of one individual by another, or duplication, enabling removal of these two samples. The remaining individuals appeared to potentially have some population substructure, but were considered otherwise unrelated, and no distinct sub-populations were identifiable. This indicates that offspring in the post-IVM sample were as individually diverse and with similar genome-wide nucleotide diversity as those in the pre-treatment sample. IVM selection pressures are therefore unlikely to occur over a large genomic region.

The pre-treatment sample was likely, due to the study design, to include individuals resistant to IVM. If IVM resistance is a dominant trait (Le Jambre et al., 2000; Sutherland et al., 2002), heterozygous adults could survive IVM treatment, with sensitive alleles retained. Therefore, the post-IVM sample could include individuals susceptible to IVM as the L3 were not directly selected by IVM. This might dilute resistance associated alleles, reducing the ability to detect regions of IVM selection. However, the lack of a distinct subset of individuals post-IVM indicates that allele frequencies of genomic regions unrelated to IVM sensitivity are unlikely to significantly change following treatment, unless linked by genetic hitch-hiking. This aspect should enhance the ability to detect markers linked to IVM resistance.

#### **4.4.2 The sample size chosen was too small**

The post-IVM sample had more private alleles than the pre-treatment sample, suggesting that the sample size was too small, or that sub-sampling of the overall population had occurred. The F-statistic,  $F_{ST}$ , was chosen to measure the degree of genetic differentiation between samples, by the degree of allele fixation between sample populations. This is commonly used in GWAS but it is

beneficial to smooth the data, reducing noise (Beissinger et al., 2015). Although smoothing was attempted within this study, RADlocus SNPs were not always smoothed efficiently. This was because, while some regions had high SNP density, other windows included a single SNP. Individual SNPs can have high  $F_{ST}$  values, hence the output was biased towards low SNP density regions, with highly differentiated SNPs. Instead it was decided to use individual SNP  $F_{ST}$  values, in addition to RADlocus haplotype diversity measures of  $F'_{ST}$  and  $D_{EST}$ . The RADlocus on Tci2 contig 146, 587481, appeared of interest in all three statistical analyses. In addition, the allele frequency change of two SNPs with high  $F_{ST}$  values within the RADlocus, with one SNP changing from roughly equal allele proportions pre-treatment to 90:10 post-IVM, was suggestive of strong positive selection at the locus. Unfortunately, following assessment of fasta sequences for each individual in Geneious v9.1.8 it became apparent that diversity increased at this locus following IVM treatment. The apparent positive selection of the SNPs appeared to be by chance, as they were present on many different genetic backgrounds, with an increased number of haplotypes present post-IVM compared with pre-treatment. Although true selection may have occurred at this locus, with multiple haplotypes carrying the allele conferring IVM resistance, this could not be concluded from these results. This also indicated that the sample size needed to be larger to detect selection by IVM using ddRAD-Seq.

#### 4.4.3 Shared RADloci were far fewer than identified RADloci

More than 3.3 M RADloci were identified by the *gstacks* module. Nevertheless, the number of informative RADloci retained comprised a tiny percentage - less than 0.5% of all 3.3 M RADloci. Shared RADloci covered less than 0.3% of the genome, and those with polymorphisms covered less than 0.1% of the genome. With even spacing of RADloci, one could expect to find an informative SNP every 5.7 kb. This would suggest sampling of the genome frequently enough to easily detect selection as the average gene length in the PRJNA72569 genome is 6145 bp (Choi et al., 2017). However, as the SNP density plots showed, SNPs are unevenly spread across the genome reducing the chance to detect narrow regions under selection. There are two main reasons for fewer shared RADloci compared to the total number of identified RADloci. Firstly, there may be differences in RADloci identified due to unequal amounts of DNA sequenced

between individuals. Secondly, different RADloci may be produced in different individuals during the digestion step. Possible causes of these two reasons will be discussed.

#### **4.4.3.1 Unequal amounts of DNA sequenced**

Individuals with a greater proportion of missing (shared) variant sites, had fewer reads aligned to the genome. This indicates that these individuals did not have fewer RADloci in common with others, which were sequenced to the same (or greater) depth, but rather were sequenced less overall.

Variation in the production of sequencing libraries between individuals could have contributed to a loss of shared RADloci. Unequal quantities of DNA were almost certainly included in the pools for sequencing, potentially due to an issue with the Qubit whilst generating the libraries. This is suggested by the variation in total sequenced reads between individuals, which correlates, in part, to the WGA batch the samples originated from. As several hundred nanograms of DNA were present in each microlitre post-WGA, errors in quantifying and pipetting would be likely to have considerable impact. The high quality of the sequenced reads suggests that differences between individuals arose prior to the sequencing step (Maroso et al., 2018). However, during sequencing unequal contributions of DNA would have led to some RADloci being sequenced in all individuals, with other RADloci missing from some individuals. This would mean that fewer than 80% of the individuals within a population would contribute to many RADloci, and these RADloci would be removed.

Differences in fragment size may also have biased selection of DNA quantity between individuals. For example, if some individuals had proportionally more DNA fragments that fell outside the 225-325 bp size range than others, then the fraction of their DNA in the pool could have been disproportionately reduced. This is unlikely to be a major factor as the population appears generally homogenous and therefore one would expect a similar size distribution in all individuals.

#### 4.4.3.2 Different RADloci may have been produced between individuals

Diversity between individuals may have both generated new RADloci and reduced shared RADloci. Over 50% of variant sites in each of the pre- and post-IVM populations were considered to be out of HWE, largely due to an excess of homozygosity which is suggestive of allele loss. This phenomenon could have occurred due to positive selection by IVM, but due to the scale, is more likely related to polymorphisms in the restriction endonuclease cut-sites, leading to allele drop out (Miller et al., 2007). Although the variant site nucleotide diversity was comparatively high (0.22), the genome-wide nucleotide diversity was only 0.002, similar to the selfing model nematode *C. elegans* (Cutter, 2006). This was much lower than expected, and probably indicates a considerable loss of variable sites, rather than a realistic measure of *T. circumcincta* population nucleotide diversity. That RAD-Seq underestimates nucleotide diversity has been recognised before (Arnold et al., 2013), due primarily to polymorphisms within the restriction enzyme cut-sites causing allele loss (Cariou et al., 2016). Equally, RADloci likely to be shared between individuals within a diverse population are perhaps more likely to be selected from genomic regions of lower nucleotide diversity.

Size selection of fragments using the Pippin Prep was efficient, based on the average insert size of reads aligned to the genome. Inefficient size selection is thus unlikely to have considerably reduced shared RADloci in this study (DaCosta and Sorenson, 2014), although it may have excluded some RADloci, or alleles, containing large structural indels within certain individuals.

Other reasons for loss of RADloci may be due to bias during WGA or PCR amplification during the generation of sequencing libraries. The multiple displacement amplification reaction (WGA) using Phi29 polymerase is much more efficient than PCR techniques. Nevertheless, amplification bias of DNA regions still occurs although considerably reduced compared with PCR (Dean et al., 2002). WGA bias could lead to a reduction in sequencing of some RADloci within an individual's genome. This artefact is exacerbated with low amounts of input DNA, with some bias noted when 0.5 ng of human DNA was used in one study, but with considerably increased bias when DNA quantity was reduced to <0.1 ng (Dean et al., 2002). A single L3 has <1 ng of DNA, and only half of this was used



as input for WGA, therefore the risk of bias was a concern. The use of trehalose within the protocol should have partially reduced this bias (Pan et al., 2008), whilst simultaneously increasing DNA yield overall (Tsai et al., 2014) providing sufficient WGA DNA for ddRAD-Seq. As a crude measure of amplification uniformity, six core eukaryotic genes were chosen and PCR amplified pre- and post-WGA, as described in 2.5.1. Results indicated that each gene was successfully amplified in each individual subsequently used for ddRAD-Seq. Despite this, synteny between species may not be conserved, such that all six chromosomes may not be represented by these genes. In addition, relative amplification was only qualitatively assessed based on band brightness, and it is possible that some regions of the genome were better amplified in some individuals than others, leading to a loss of shared RADloci.

Finally, it is also possible, due to the use of a fragmented, partially assembled genome, that haplotypic diversity within the *Tci2* genome has not yet been resolved. If so, allelic reads of individuals could be falsely split across multiple RADloci. This is difficult to detect without gene annotations. However, it is another potential explanation for the high number of invariant RADloci, the large number of variant sites out of HWE and the low percentage of RADloci shared between individuals.

The number of RADloci retained reduced considerably if they were required to have been sequenced in all individuals. Conversely, more RADloci were retained if fewer than 80% of individuals were required. Less than 0.03% of the 685 MB *Tci2* genome was represented across all 75,517 RADloci, and less than 0.01% across those containing variant sites. Using the same sequenced data, it could be beneficial to sample a greater proportion of the genome, by allowing retained RADloci to be shared by fewer individuals. Relative differences in  $F_{ST}$  values are recognised when differing numbers of individuals are included in an analysis (Choi et al., 2017). If the number of individuals with shared RADloci was reduced, this would increase the variation in the number of individuals per RADlocus (e.g. some RADloci would have only 60% of individuals, others would be sequenced in all individuals). This would necessitate multiple separate analyses, which might not be directly comparable. Furthermore, given the diversity of the population and the low sample size to begin with, this did not seem appropriate.

Therefore, only RADloci present in 80% or more of individuals were examined in this study.

#### 4.4.4 Linkage disequilibrium decayed rapidly

Identifying markers of IVM resistance requires either the marker to be causative, or to be consistently linked to a causative variant. LD was found to rapidly decay, such that the distance to the half maximal  $r^2$  value was always the minimum distance between variant sites tested. To try to obtain a more meaningful estimate, the distance to the mean intercontig LD was obtained. This was calculated to be only 2805 bp. To put this in context, the average gene size in the *T. circumcineta* PRJNA72569 genome was 6415 bp (Choi et al., 2017). When the SNP density per kb was extracted and plotted it was clear that RADloci retained were unevenly spread across the genome, with large regions with no variant sites within them. This could greatly reduce the chance of identifying markers of IVM resistance in such an outbred population, where the genomic footprint of a resistance locus might be small (Smith and Haigh, 1974). Recent work in *H. contortus* laboratory and field isolates has identified a large locus under IVM selection, in a region of low recombination (Redman et al., 2012; Rezansoff et al., 2016; Doyle et al., 2019; Sallé et al., 2019), which would be likely to be detected using a RAD-Seq approach. However, it would appear the architecture of resistance may differ in field populations of *T. circumcineta*. Alternatively the IVM locus may be in a highly fragmented region of the assembly, or was simply not sampled in this study.

It is known that recombination rates and nucleotide diversity vary across a genome, and even along a chromosome (Doyle et al., 2018). This could allow some regions of low recombination, and high LD, to prove useful, even with low SNP density (McKinney et al., 2017). Furthermore, if IVM resistance had been recently selected within a population, which might have occurred on this farm based on the FECRT history, larger regions of low diversity and high LD might be present surrounding IVM resistance loci. This could allow the detection of IVM resistance markers even with low LD genome-wide (Catchen et al., 2017; Lowry et al., 2017). To determine whether any contigs with highly differentiated RADloci also had high LD, contigs containing RADloci with high  $F_{ST}$  variants were extracted. Despite the presence of SNPs of high  $F_{ST}$ , suggesting potential

selection pressure by IVM, LD decay appeared similar to genome-wide estimates. No association between mean contig  $F_{ST}$  and mean contig LD was present in the study data. Although interpretation of this result is limited by variation in contig lengths (longer contigs may have short regions of both high LD and high  $F_{ST}$ ), and SNP density, it lends support to the theory that genetic footprints of IVM resistance are narrow, or that the genetic basis has arisen, or is currently located, on multiple genetic backgrounds.

## 4.5 Conclusion

Overall, this ddRAD approach was of limited value for the detection of IVM resistance loci in a UK field population of *T. circumcincta*. This was due to a lack of shared data between individuals and an insufficient SNP density across the genome to compensate for low LD. However, additional limitations were found in the use of the Stacks v2.0b program which, at the time of writing, has been updated nine times since it was used for this study. Nevertheless, from this analysis we were able to determine the need to sample more individuals to attain a better estimate of allele frequency within a population. We could also calculate the approximate LD and we obtained haplotypic sequencing data from a considerable number of individuals which could be useful later.

As the *T. circumcincta* genome assembly improves, this analysis, using an updated version of Stacks, or similar software/pipelines, could be repeated and may provide improved outcomes. With a more assembled reference genome, ddRAD-Seq or similar techniques may become more useful. The large genome size still prevents WGS of individuals, particularly when high levels of genetic diversity necessitate large numbers of individuals for sufficient power. Instead, other options are available, including pooled sequencing of individuals, allowing both a larger sample size, and an increased fractional coverage of the genome, whilst maintaining comparatively low sequencing costs. Based on the results of this study, WGS with a larger sample size may be a more appropriate approach to study IVM resistance in this species.

## 5 A tale of two farms: validation of a Pool-Seq technique to investigate IVM resistance in UK field populations

### 5.1 Abstract

The IVM locus associated with resistance may be small in UK field populations of *Teladorsagia circumcincta*. If IVM resistance is dominantly inherited, diversity of, or around the specific genes or variants under selection by IVM may still be comparatively high even following direct selection with IVM treatment. High individual diversity can complicate identification of suitable markers with small sample sizes. However, sequencing sufficient numbers of worms individually is expensive. Using a whole genome re-sequencing approach, I sequenced pools of 91 larvae obtained from adults treated with IVM, pre- and post-IVM treatment. Two geographically separated UK farms were used and were treated independently throughout the study. Despite diversity within the population, it was assumed that alleles conferring, or associated with IVM resistance would be enriched in the post-IVM samples. In contrast, neutral regions of the genome were likely to be similar between pre- and post-IVM sample populations. Population genetics analysis identified multiple regions of the genome to be potentially under selection by IVM. To identify whether a common genetic basis between farms exists, I included two UK farms in this genome wide association study. Genetically differentiated regions were predominantly specific to each individual farm; however, sixteen 10 kb windows were identified to be potentially under IVM selection on both farms. No single large locus was identified to be under IVM selection.

### 5.2 Introduction

Whole genome re-sequencing (WGS) of pooled individuals (Pool-Seq) can provide a highly cost-effective approach for population genomic analysis. Pool-Seq has been used to detect population substructure (Phair et al., 2019), changes in populations over time (Phillips et al., 2016), somatic variants in cancer studies (Ding et al., 2012) and in genome wide association studies (GWAS), linking a variant to a phenotype of interest (Ehrenreich et al., 2010). When sequencing individuals separately, many sequencing reads used to determine a genotype at

a genomic site are surplus to requirements. This reduces the cost-benefit of the sequencing performed. In contrast, using Pool-Seq, most individuals within the pool are usually sequenced less than once, such that each read contributes to the overall estimation of allele frequency (Futschik and Schlotterer, 2010). For example, using a model for diploid individuals, Gautier et al. (2013a) found that for half the sequencing cost, twice the number of individuals could be pooled providing a more accurate estimation of the population allele frequency compared with sequencing individuals separately.

WGS also allows for much wider sequencing coverage of the genome than reduced representation library methods (e.g. ddRAD-Seq, exome sequencing, RNA-Seq). This provides several benefits, including: the increased ability to identify small footprints of genetic selection, to obtain a more accurate estimate of the nucleotide diversity, to identify the presence of indels and transposable elements and to include low complexity regions within the sequenced results. In an outbred population, with high nucleotide diversity and rapid decay of LD, as was found in this study, WGS can therefore provide many benefits over reduced representation methods for a GWAS. For example, Barnes et al. (2017) were able to confirm a hard sweep around the *CYP6P9a* gene, in populations of *Anopheles funestus* using Pool-Seq. This corresponded to intense use of pyrethroids since 2002, such that the 2002 and 2014 samples displayed high genetic differentiation at this locus. Cheeseman et al. (2015) used Pool-Seq to confirm the presence of a soft sweep in *Plasmodium falciparum*. Multiple haplotypes were selected around the *kelch* gene, associated with artemisinin resistance in this species. Although the authors were able to detect the locus, using  $F_{ST}$ , various other neighbouring loci had higher  $F_{ST}$  values - a factor which may be important in the present study. In contrast, using microsatellites, Redman et al. (2012) identified that the marker, Hc\_8a20, was linked with *H. contortus* IVM resistance phenotypes, but were unable, particularly with an incomplete genome assembly, to identify which genes were associated with IVM resistance. More recently, using WGS, a region containing many genes on Chromosome V has since been identified (Doyle et al., 2019).

The benefit of Pool-Seq to determine genome-wide nucleotide diversity allows for calculation of Tajima's  $\pi$  and Tajima's  $D$  for each site or region (Kofler et

al., 2011b). These statistics provide a measure of diversity for a population, which can provide evidence of fixation of alleles due to a reduction in diversity over that region (Tajima, 1989). This measure can therefore be useful to confirm regions under selection (Doyle et al., 2019), or to identify regions of fixation.

As found in the ddRAD-Seq experiment, the use of a selective reduced representation technique was not the best option for the highly polymorphic *T. circumcincta* field population. Rapid decay of LD also reduced the ability to detect regions of selection across the genomes within the population. It was decided to perform WGS to sequence the entire genome, using a larger sample to improve the accuracy of the population allele frequencies obtained. To perform WGS of individuals was prohibitively expensive. Instead, Pool-Seq would enable an increased sample size, with a better allele frequency determination for the same budget. In addition, the amount of DNA required for Pool-Seq is less than for ddRAD-Seq. This allows gDNA of pooled L3 to be used directly, without the need for whole genome amplification, reducing library preparation costs. In addition, by requiring fewer library preparation steps prior to pooling of individuals, variation between individuals within and between pools is likely to be reduced. If IVM resistance is dominant as suggested in adult *T. circumcincta* (Sutherland et al., 2002) and in *H. contortus* (Le Jambre et al., 2000), or partially dominant as suggested in L3 of *T. circumcincta* (Sutherland et al., 2003b), then survival of heterozygous adults may occur following IVM treatment. Offspring used in this study could therefore contain heterozygous or homozygous susceptible individuals. These haplotypes and genotypes could dilute signals of IVM selection, although the region under selection should still be visible. If, however, SNPs or haplotypes related to IVM resistance reach fixation, the selection of the farm population used will be critical. Regions of fixation within a population can have very low minor allele frequencies, which also produce low  $F_{ST}$  values and makes identifying such regions difficult as, both before and after treatment, most individuals will be homozygous for the variant.

Pooled sequencing can be performed using barcodes to identify and separate individuals or groups post-sequencing, potentially providing additional useful data (Wong et al., 2013), however this can rapidly increase library preparation and labour costs. Alternatively, sequencing can be performed as a mixed pool

(Pool-Seq), with no ability to separate and identify individual DNA sequences post-sequencing, which is considerably cheaper, but brings with it specific concerns as to unknown bias within the data, which cannot always be accounted for (Futschik and Schlotterer, 2010; Gautier et al., 2013a). A primary concern for many researchers is that of biased allele frequencies, both in terms of actual allele frequencies within the sample, and in degrees of variation between samples. Bias could occur following pooling of unequal amounts of input DNA from individuals within the pool (Zhu et al., 2012), or variation in sequencing coverage between individuals at a locus (Anderson et al., 2014). PCR duplication, if PCR was used in library preparation, can bias allele frequencies and estimates of the effective population size (Futschik and Schlotterer, 2010). Other, more generic sequencing errors, which affect both individual and pooled approaches, include either dependent (repeatedly the same nucleotide substituted, e.g. A, A, A instead of T), or independent (different nucleotides substituted, e.g. A, G, C instead of T) errors which can confound SNP calling, particularly in pooled samples if the minor allele read count is low (Futschik and Schlotterer, 2010).

The power of a sequencing study is its ability to successfully detect true positive variants (low Type II error). In other words, a highly powerful study has a low chance of accepting the null hypothesis of no difference when it should, in fact, be rejected. When sequencing individuals, raising the coverage provided can allow an increase in power to detect true SNPs and also determine false positives, especially if errors are independent, although this is less effective for dependent errors (Futschik and Schlotterer, 2010). Various simulations and real-life experiments have determined that Pool-Seq can easily detect such variants, as well as selective sweeps, often at coverage considerably lower than that required for individuals. For example, Boitard et al. (2012) suggest, from simulations, that coverage of 0.5x per chromosome is sufficient to determine the allele frequency spectrum in a pool of 200 chromosomes. They were able to detect known regions of selection along the X-chromosome using pools of *Drosophila melanogaster*, sequenced at approximately 1x per individual (Boitard et al., 2012). Several guidelines are important to follow however: firstly, to ensure that sufficient individuals are sequenced within the pool, with each individual contributing equal amounts of DNA (Zhu et al., 2012; Gautier et al.,

2013a; Anderson et al., 2014; Rode et al., 2018). Minimum cut-offs for the minor allele count should reflect the expected sequencing error and coverage depth (Kofler et al., 2011a). The library preparation technique appears less important than the coverage (Kofler et al., 2016b). Increasing coverage of pools containing 554 to 736 *Drosophila* from 30 to 50x, improved estimates of allele frequencies within a population, with differences in allele frequency estimation reduced by 27% (a difference in frequency estimate of 0.016) (Kofler et al., 2016b). Despite this, the degree of variation in coverage across a genome was not found to greatly affect results, providing the model used accounted for this (Boitard et al., 2012). Lastly, using technical replicates can greatly improve confidence in results (Robasky et al., 2014).

Again using progeny of adults treated with IVM as part of a faecal egg count reduction test (FECRT), I used Pool-Seq to identify regions of the genome potentially under selection by IVM. In addition to Farm 2 (also used for ddRAD-Seq) I also selected an additional UK farm (Farm 3) for the Pool-Seq study. Farm 2 and Farm 3 were geographically separate populations. They had different IVM FECR *T. circumcincta* percentages (comparatively high efficacy on Farm 2, low efficacy on Farm 3). Each farm was analysed separately - identifying regions of genetic differentiation between the pre- and post-treatment populations. Following within farm analysis, results of loci potentially under selection by IVM on each farm were also compared based on the hypothesis of a common locus conferring IVM resistance. Results describe the validation and outcome of this technique on these two farms.

## 5.3 Results

### 5.3.1 Farms used and FECRT data

Two UK farms were chosen for the analysis, referred to here as Farms 2 and 3. Farm 2 has been previously described in 4.3.1 and a summary of the FECRT is repeated in Table 5.1 for comparison with Farm 3.

Farm 3 is in southeast Scotland and in July 2015 a FECRT was performed by colleagues at the Moredun Research Institute (MRI). Anthelmintic efficacies were -34.4%, 96.0% and 22.4% for BZ, LEV and IVM respectively (Table 5.2). Pre-



treatment egg counts were 135.8 epg and 181.0 epg for BZ and IVM treatment groups respectively. All L3 strongyles provided by the MRI and speciated by PCR pre- and post-IVM were *T. circumcincta*, and all but one post-BZ L3 were also *T. circumcincta*. A further two groups of lambs were used to test the combined efficacy of co-administering BZ and IVM, and LEV and IVM. These demonstrated improved anthelmintic efficacies compared with BZ or IVM given as a single active alone, reducing egg output by 54.5% and 89.1% respectively (Morrison, *pers comm*, Table 5.2).

**Table 5.1: Faecal egg count reduction test results on Farm 2, 2013 and 2016.**

Anthelmintic	FECRT 2013 (%)	FECRT 2016 (%)	FECRT 2016 efficacy against <i>Teladorsagia circumcincta</i> (%)
BZ	83.7%	87.7%	ND
LEV	92.5%	95.4%	ND
IVM	93.0%	87.4%	81.9%
MOX	ND	98.6%	ND

Overall percentage reduction in strongyle egg output following anthelmintic treatment is shown, with the proportional reduction percentage of *Teladorsagia circumcincta* indicated for those samples where L3 larvae were speciated by PCR of the ITS2 region. ND = Not done.

**Table 5.2: FECRT results on Farm 3, 2015.**

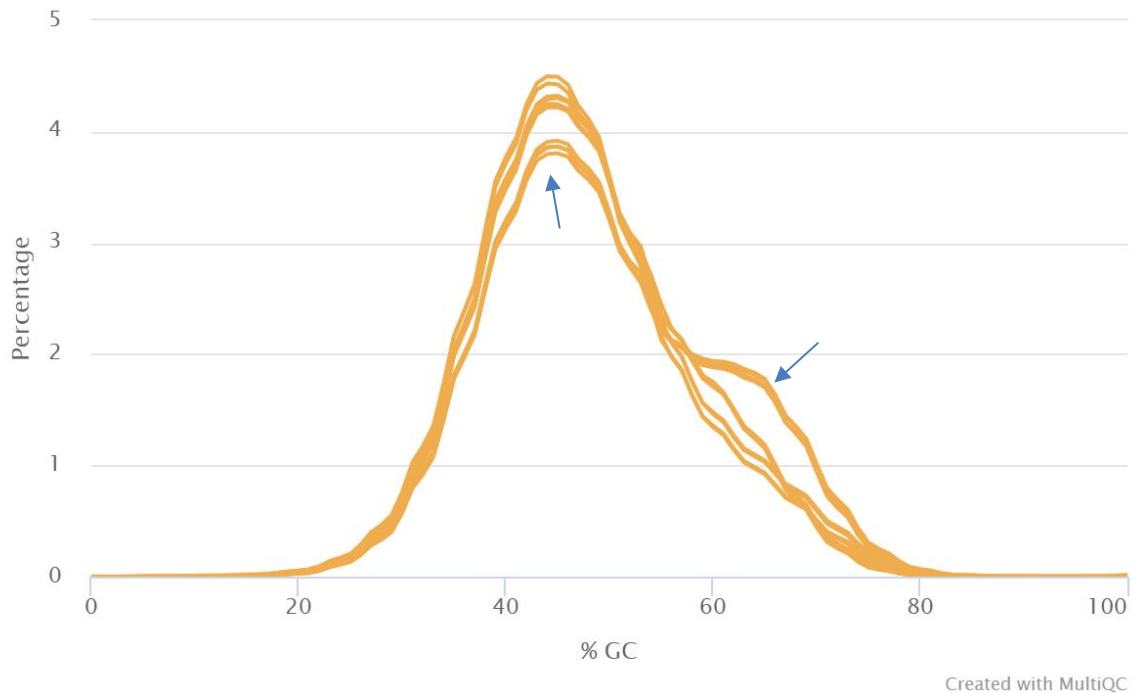
Anthelmintic	FECRT 2015 (%)	FECRT 2015 efficacy against <i>Teladorsagia circumcincta</i> (%)
BZ	-34.4%	-32.5%
LEV	96.0%	ND
IVM	22.4%	22.4%
BZ and IVM	54.5%	ND
LEV and IVM	89.1%	ND

Overall percentage reduction in strongyle egg output following anthelmintic treatment is shown, with the proportional reduction percentage of *Teladorsagia circumcincta* indicated for those samples where L3 larvae were speciated by PCR of the ITS2 region. ND = Not done.

### 5.3.2 A smaller percentage of reads aligned post-treatment compared with pre-treatment

Read quality was good, with an average Phred score of greater than 38 for the first read of each pair and an average score of greater than 30 (99.9% chance that the base called is correct), for the second read in each pair. Post-IVM samples on Farm 3 appeared to have a subset of reads with higher GC content compared with other samples (see arrows, Figure 5.1). This may have been due to bacterial contamination, with a higher GC content (Barbu et al., 1956).

Results of a Kraken report (Wood and Salzberg, 2014) produced by the WSI suggested approximately 10% of reads were proteobacteria for each sample sequenced. However, these results did not provide evidence of species or abundance differences in the post-IVM samples on Farm 3 to explain the GC results. Average GC content varied between 47 and 49% (Table 5.3), similar to that expected for *T. circumcincta* (Choi et al., 2017). Read duplication was reasonable with, at most, 14.5% of reads duplicated.



**Figure 5.1: Pool-Seq read GC content distribution. Plotted by MultiQC (Ewels et al., 2016), orange lines indicate the GC percentage content distribution as a percentage of all reads sequenced for a sample. Each sample has two lines, due to paired-end sequencing. Blue arrows indicate the Farm 3 post-IVM samples.**

Between 60.6 and 70.1% of reads aligned to the genome (Table 5.3), with fewer reads mapping in their pair to the genome (48.5 to 56.4%). The majority of reads which were mapped, but not properly paired, were included in the number of reads with a mate mapping, with good quality score, to a different scaffold, which reflected the fragmented nature of the reference genome. As for ddRAD-Seq a lower percentage of reads in the post-IVM samples mapped to the genome compared with the pre-IVM samples.

**Table 5.3: Summary of Pool-Seq alignment.**

Sample Name	Reads Mapped (M)	Reads Mapped (%)	Reads 'properly paired' in mapping <sup>1</sup> (%)	Pre-alignment mean GC content (%)
Farm 2 Pre-treatment	451.8 M	70.1%	56.4%	47%
Farm 2 Post-IVM	437.4 M	66.0%	52.6%	47%
Farm 3 Pre-IVM replicate A	470.9 M	68.4%	54.5%	47%
Farm 3 Pre-IVM replicate B	460.3 M	68.7%	54.8%	47%
Farm 3 Post-IVM replicate A	418.3 M	60.6%	48.5%	49%
Farm 3 Post-IVM replicate B	422.9 M	61.6%	49.4%	49%

Reads aligned to the reference genome for each Pool-Seq sample are shown. <sup>1</sup>Reads which are 'properly paired' are those which have been mapped in their pair, rather than with the pair split over a considerable distance, or to a separate contig.

In total, between 17.7 and 17.8% of the total number of aligned reads had a MAPQ=0 for each sample. The mapping error rate output by SAMtools *stats* was similar across all samples ranging from 2.94E-02 to 2.99E-02 (Li et al., 2009).

### 5.3.3 Coverage varies between pre- and post-treatment samples

The average mapped coverage across the genome as a whole,

$$\left( \frac{\text{number of reads} \times 75 \text{ bp read length}}{\text{genome length}} \right),$$

varied from 45.7x to 51.5x (Table 5.4). The coverage frequency calculated by SAMtools v1.3 *stats*, revealed that pre-treatment samples tended to have a greater number of sites with higher coverage depths of 60x or more compared with post-treatment samples, which had a greater proportion of sites with coverage depth <40x (Figure 5.2). This reflected the total number of reads aligned to the reference genome, with more reads aligned per pre-treatment sample than post-treatment sample (Table 5.4). Most genomic sites had a coverage of 100x or less. Bedtools v2.17.0 *multicov* (Quinlan and Hall, 2010; Quinlan, 2014) was used to calculate the coverage in distinct windows across the genome, equivalent to the average gene size in the PRJNA72569 *T. circumcincta* genome (Choi et al., 2017). Each sample had extremely similar patterns of coverage across the genome. Data from the Farm 2 pre-treatment sample is

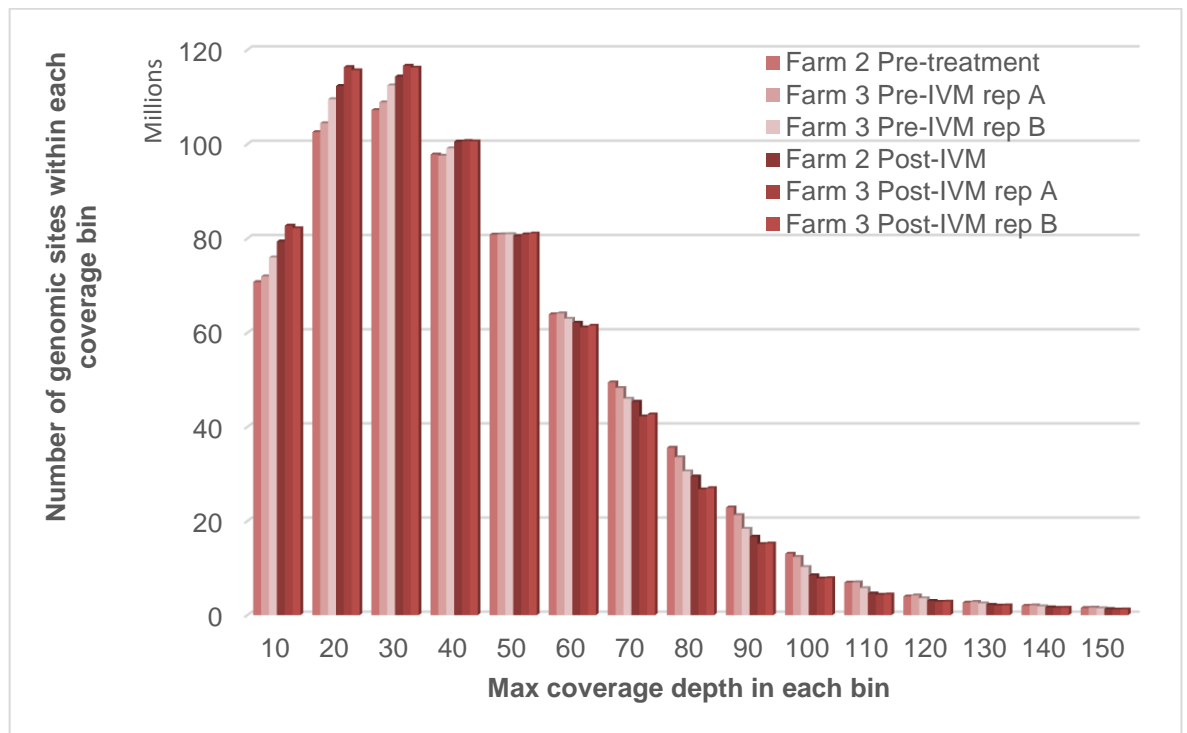
provided as an example, highlighting the variability in coverage along the genome (Figure 5.3). The mean mapped depth per site, as calculated by SAMtools v1.3 *depth*, filtering for mean BQ of >20 was lower than the average coverage per site calculated, ranging from 35.5x to 39.2x (Table 5.4).

**Table 5.4: Pool-Seq genome-wide mapped coverage per sample.**

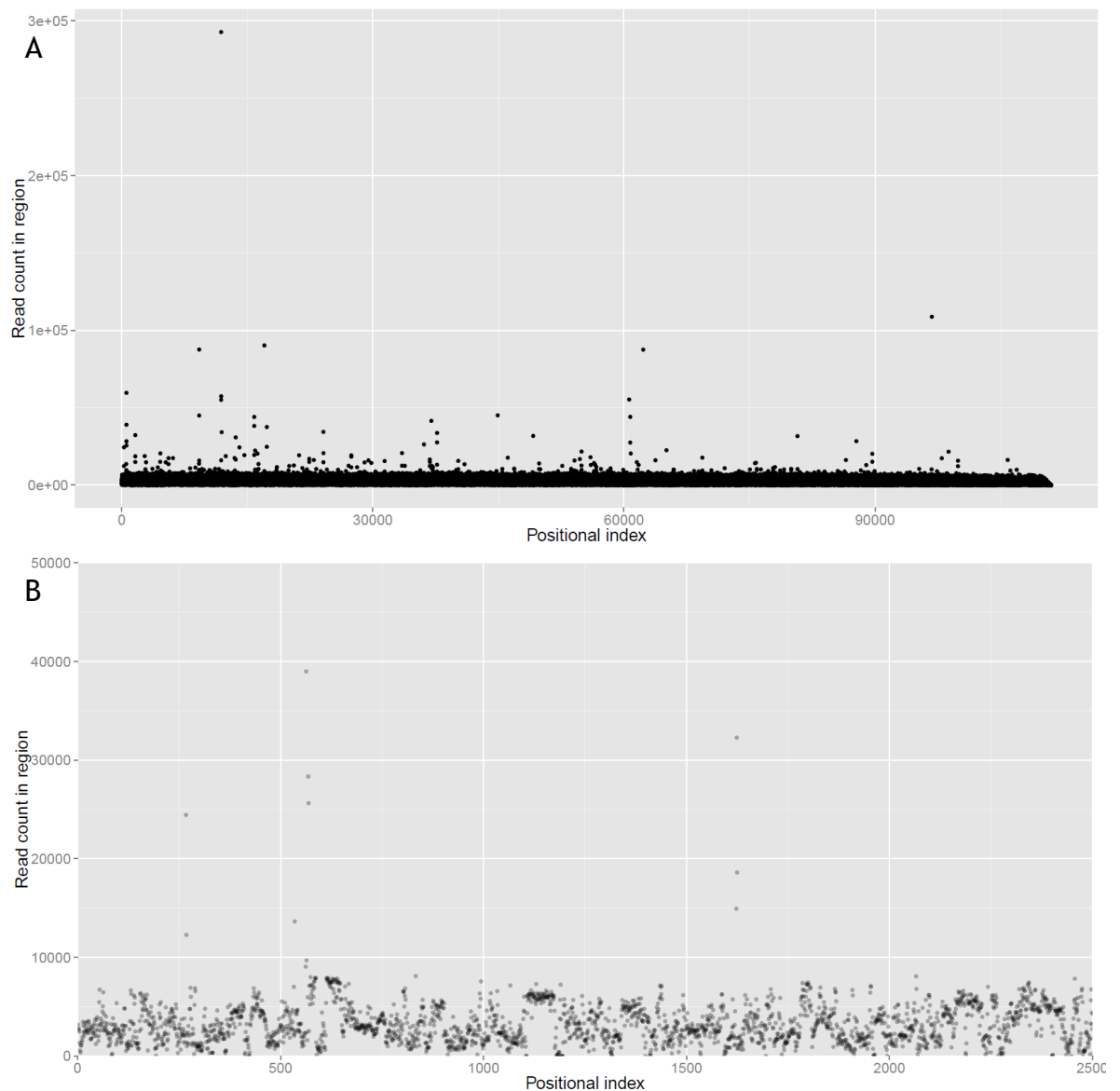
Sample Name	Reads mapped (M)	Mapped coverage across the genome as a whole (685.9 Mb genome) <sup>1</sup>	Mean mapped depth per site <sup>2</sup>
Farm 2 Pre-treatment	451.8 M	49.4x	39.2x
Farm 2 Post-IVM	437.4 M	47.8x	36.3x
Farm 3 Pre-IVM replicate A	470.9 M	51.5x	38.9x
Farm 3 Pre-IVM replicate B	460.3 M	50.3x	37.5x
Farm 3 Post-IVM replicate A	418.3 M	45.7x	35.5x
Farm 3 Post-IVM replicate B	422.9 M	46.2x	35.7x

<sup>1</sup>The coverage was calculated by the following equation:  $((N \text{ reads in millions} * 75) / 685.9)$ .

<sup>2</sup>Note the depth per site was calculated by SAMtools, and *awk* was used to extract the mean depth per site for each sample.



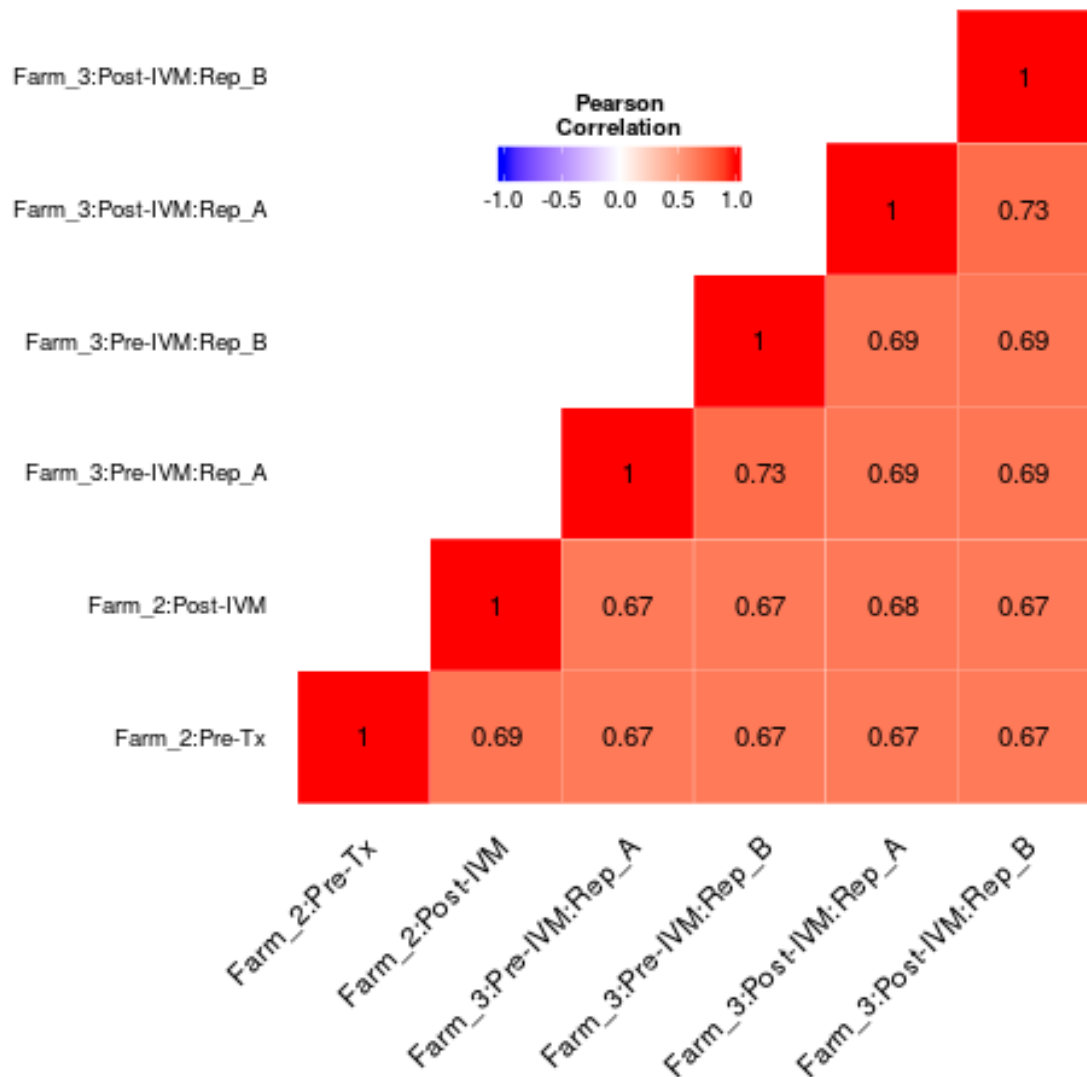
**Figure 5.2: Coverage depth site frequency.** All genomic sites were binned by coverage depth, using bins of coverage range ten, i.e. the first bin has sites with coverage depth 1-10, the next 11-20 etc. Pre-treatment samples are coloured lighter (and to the left) than post-treatment samples (to the right). Only bins of up to 150x coverage depth are shown.



**Figure 5.3: Read count in 6145 bp windows along the genome. Data plotted are for Farm 2 pre-treatment. A. All windows are plotted sequentially, from contig 1 on the left, to contig 8025 on the right, which is roughly in contig size order from longest to shortest. B. Enlarged plot of the first 2500 windows, showing only those with read count up to 50,000.**

### 5.3.4 Sample pools are diverse, with some evidence of selection by IVM

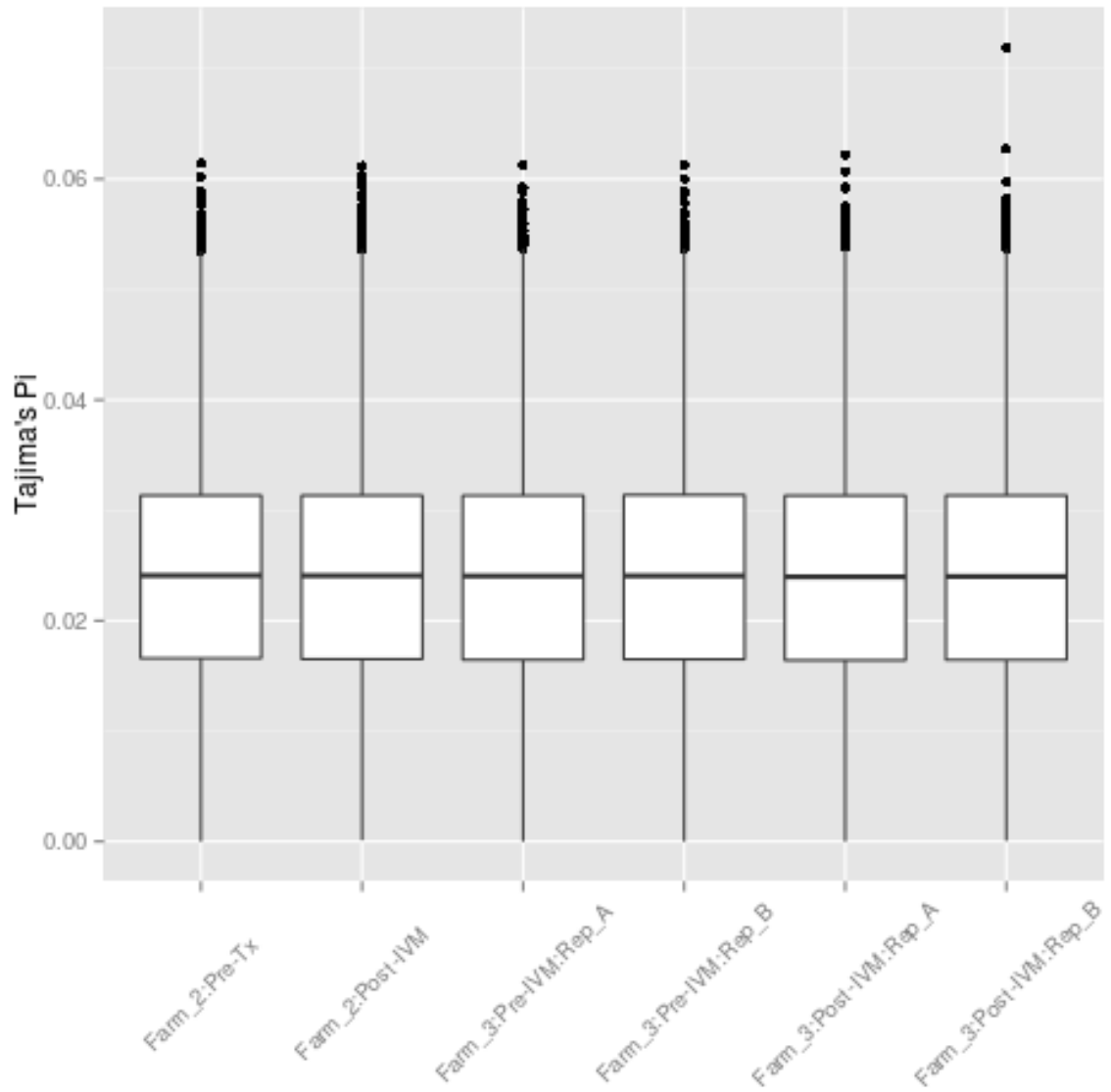
Biallelic SNPs were used to assess sample relatedness. Pearson correlation coefficients were calculated for the major allele frequencies at each SNP, and samples were found to be diverse. Although technical replicates were slightly more similar to one another ( $r = 0.73$ ), correlation between samples varied between 0.67 and 0.69 (Figure 5.4). There did not appear to be any grouping of samples by farm or by treatment time point.



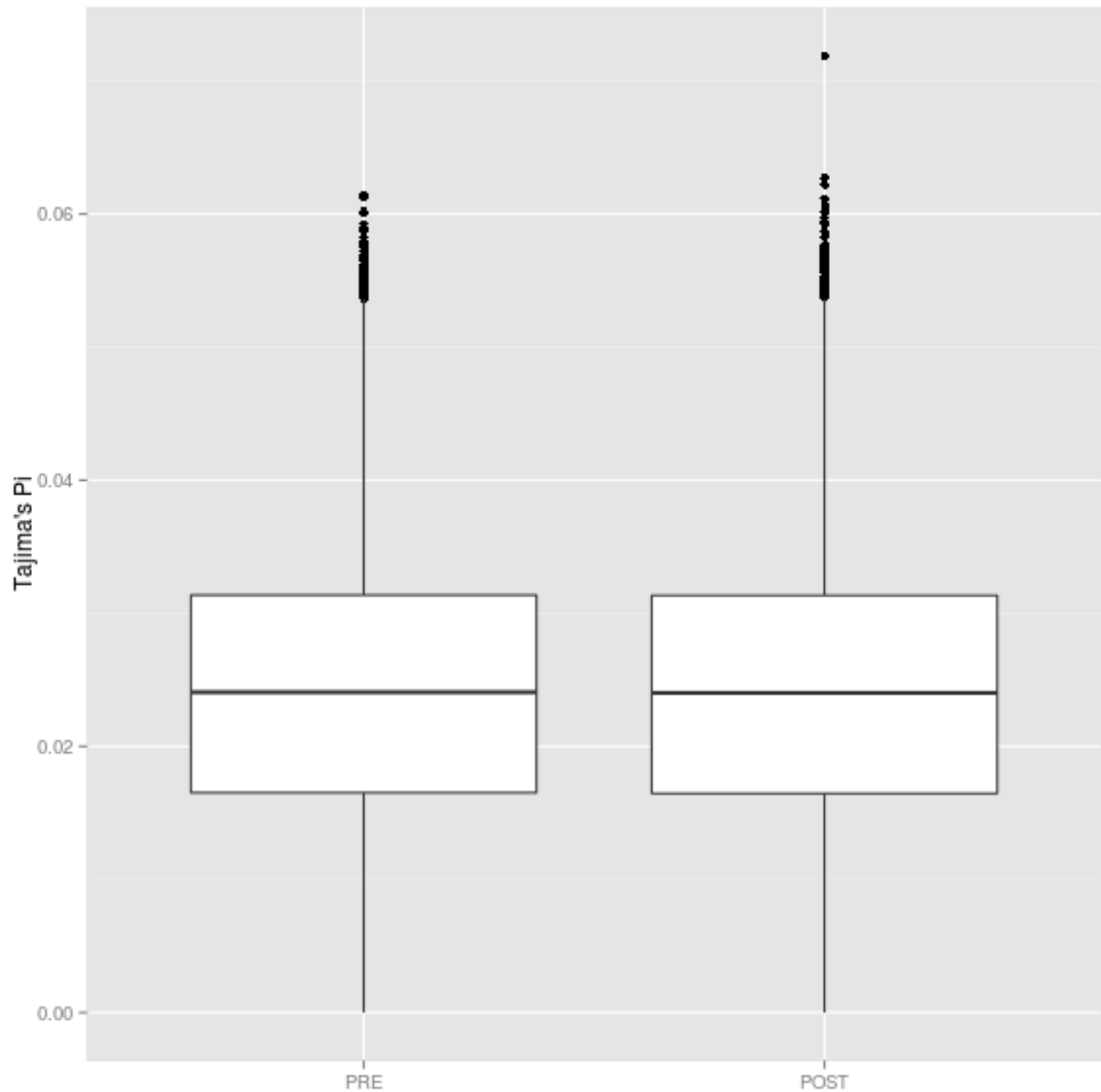
**Figure 5.4: Pearson correlation coefficients of genome-wide allele frequencies. Biallelic SNPs were used to calculate correlations between samples. A heatmap is shown, with positive correlations in red, and the correlation coefficient given for each pairwise comparison.**

Tajima's  $\pi$  (the nucleotide diversity, ( $\pi$ )) was calculated for each sample using PoPoolation v1.2.2 (Tajima, 1989; Kofler et al., 2011a) using distinct 10 kb windows. Tajima's  $\pi$  is a measure of the number of pairwise differences between sequences. The greater the number of pairwise differences, the higher the value of  $\pi$ . However, Tajima's  $\pi$  has a tendency to underestimate the presence of rare mutations, and will be influenced by the alignment of reads to the reference genome. Genome-wide Tajima's  $\pi$  was similar across all sample populations, averaging 0.024 for each sample tested (0.01 standard deviation). This would suggest a 2.4% chance of a mutation at any given site within the pool of individuals. Median values of nucleotide diversity were almost the same as the

mean values for each sample. Samples were directly compared using a Wilcoxon rank sum test in R v3.1.2, with continuity correction. No significant differences in genome-wide nucleotide diversity were identified between samples (Figure 5.5), or by treatment status (Figure 5.6).



**Figure 5.5: Tajima's Pi by sample. A boxplot of the genome-wide Tajima's Pi values is shown.**



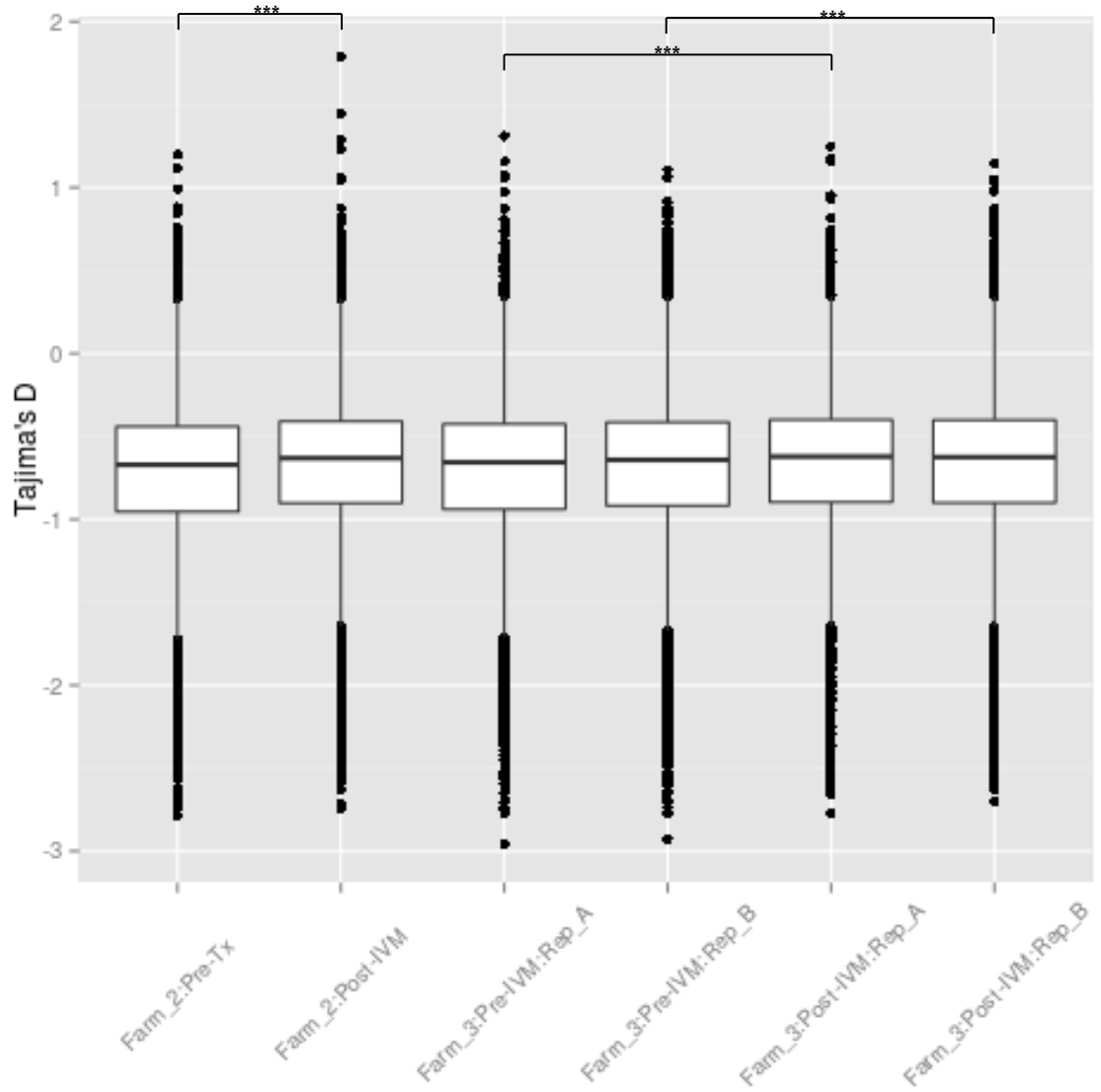
**Figure 5.6: Tajima's Pi by treatment time point. A boxplot of the genome-wide Tajima's Pi values is shown. Samples are grouped by treatment time point.**

Tajima's D is a measure estimating the difference between the number of total segregating sites and the number of pairwise differences between DNA sequences (Tajima, 1989). Directly following a selection event, dependent on the severity of such an event, there may be very few rare alleles and Tajima's D may be positive. Over time, as new mutations occur, rare alleles will increase and Tajima's D will become negative. As the population breeds, alleles will be redistributed becoming more common, and Tajima's D values will return towards zero.

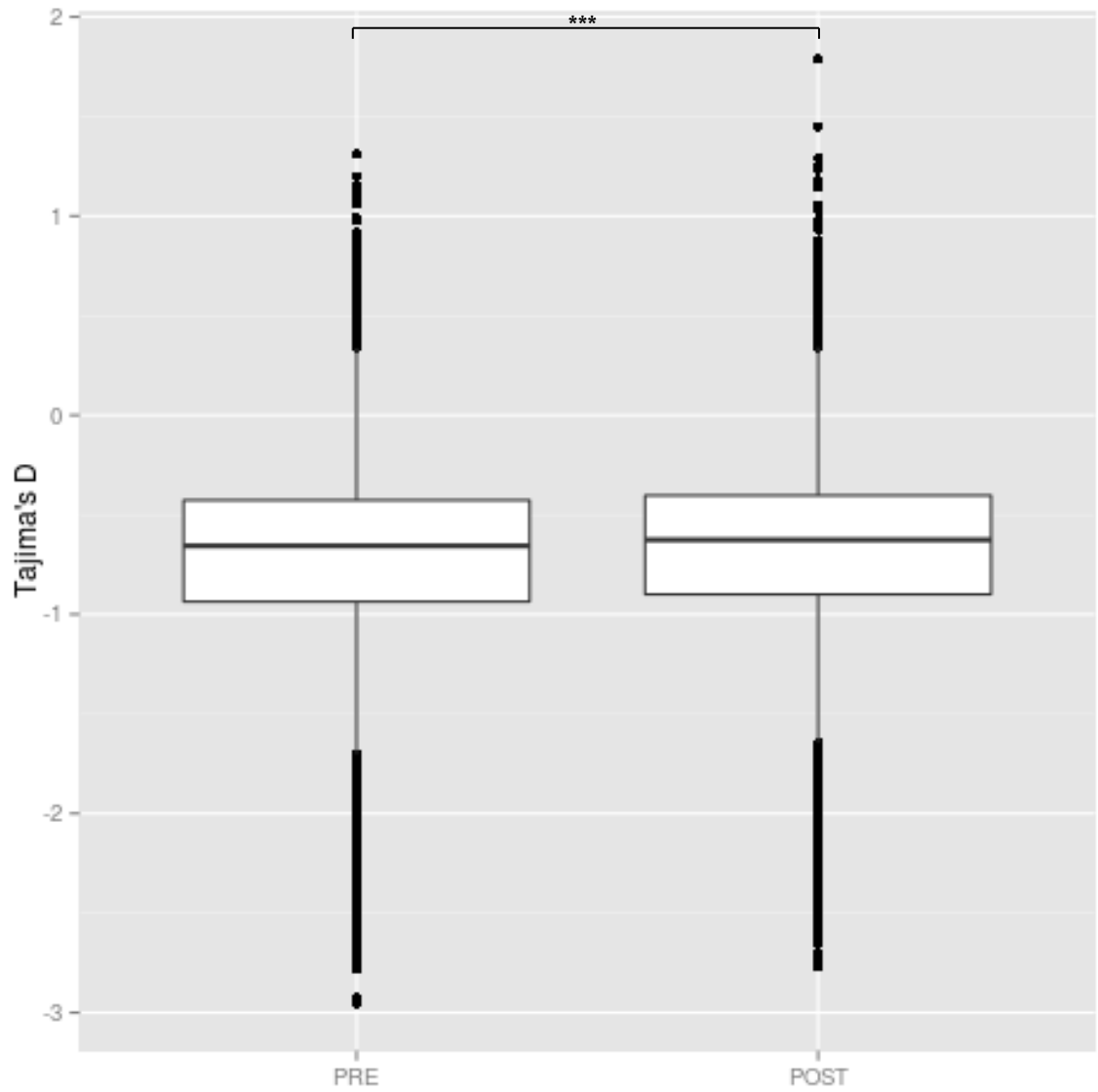
For each sample population, PoPoolation v1.2.2 was used to calculate Tajima's D using distinct 10 kb windows along the genome. The mean value for all samples



was close to -0.7, with the median value for each slightly more positive, ranging from -0.62 to -0.67 (Figure 5.7). Tajima's D varied considerably across the genome for each sample (Figure 5.9 and Figure 5.10). Interestingly, all post-IVM samples were slightly, but significantly, less negative compared to the pre-treatment samples ( $p < 0.001$ , Wilcoxon sum rank test, Figure 5.7 and Figure 5.8), suggesting fewer rare alleles within these samples. Most regions had Tajima's D values of between 0 and -1, indicating that any selection pressure on the populations has not occurred genome-wide, as usually only values of greater or less than 2 and -2 would be considered to suggest significant deviation (Tajima, 1989). Although several regions had values between -2 and -3 (Figure 5.9 and Figure 5.10), these were difficult to assess visually. They may represent just one or two fragmented loci, or many separate loci. Moderately negative regions tended to be the same between technical replicates (Figure 5.10, A & B), although some differences were apparent, particularly on shorter contigs. If highly negative regions were the same in both pre- and post-treatment samples (highlighted in colour, Figure 5.9 and Figure 5.10), post-IVM samples were less negative than pre-treatment samples. One region stood out on Farm 2 as having a greater number of highly negative post-IVM windows, with few pre-IVM windows in the same region (near positional index 25,000). However, there were still a few windows in this region with highly negative values in the pre-treatment sample, and due to the scale it is difficult to confirm whether these points represent a single region, or many. A similarly positioned window was also present on Farm 3 but was more negative in the pre-IVM samples than the post-IVM samples.



**Figure 5.7: Tajima's D by samples.** A boxplot of the genome-wide Tajima's D values is shown. Samples were compared using a Wilcoxon sum rank test in R as shown. The medians of all pairwise comparisons indicated (\*\*\*) were significantly different  $p < 0.001$ .



**Figure 5.8: Tajima's D by treatment time point.** A boxplot of the genome-wide Tajima's D values is shown. Samples are grouped by treatment time point. The medians are significantly different ( $p < 0.001$ ).

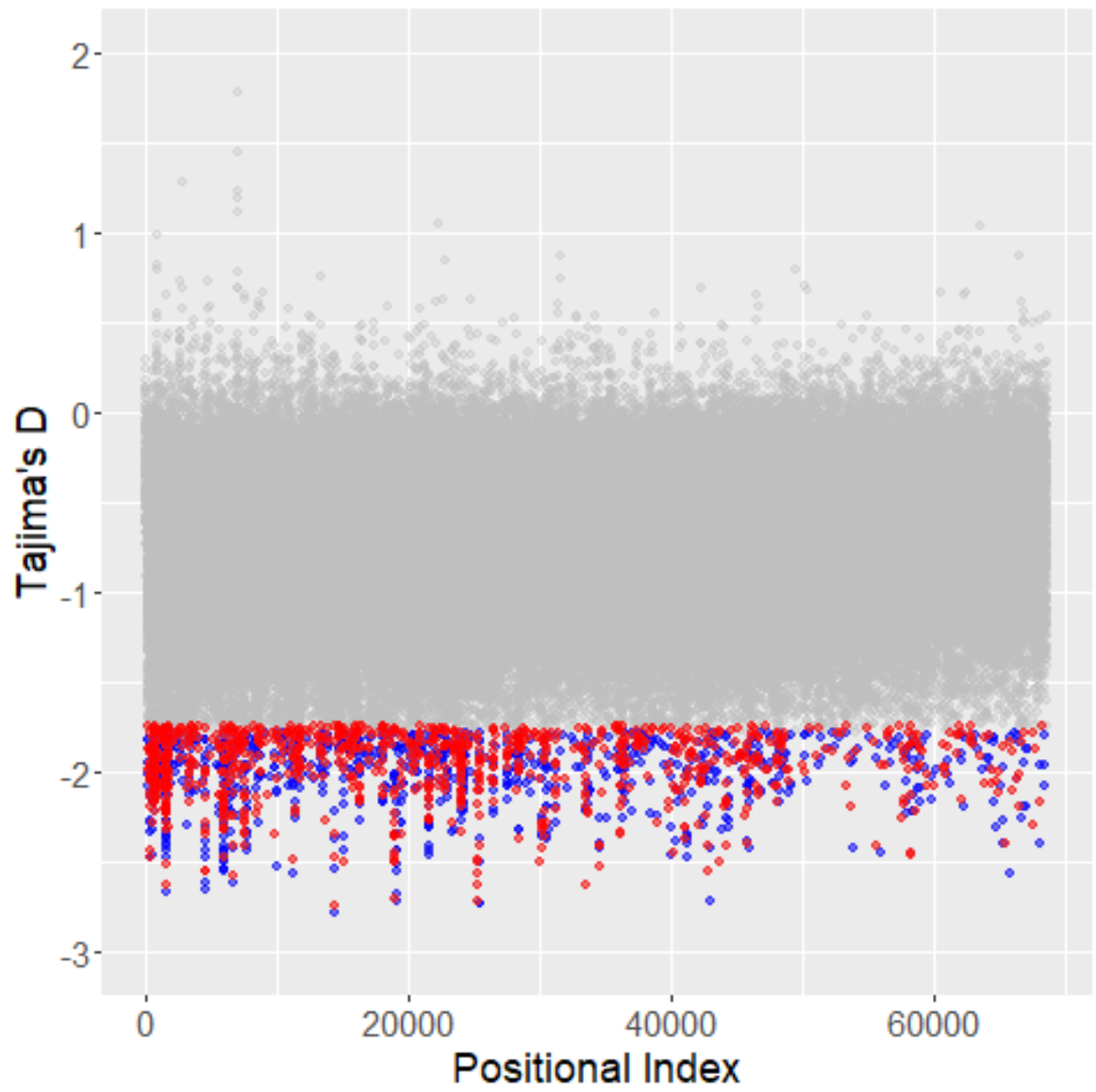
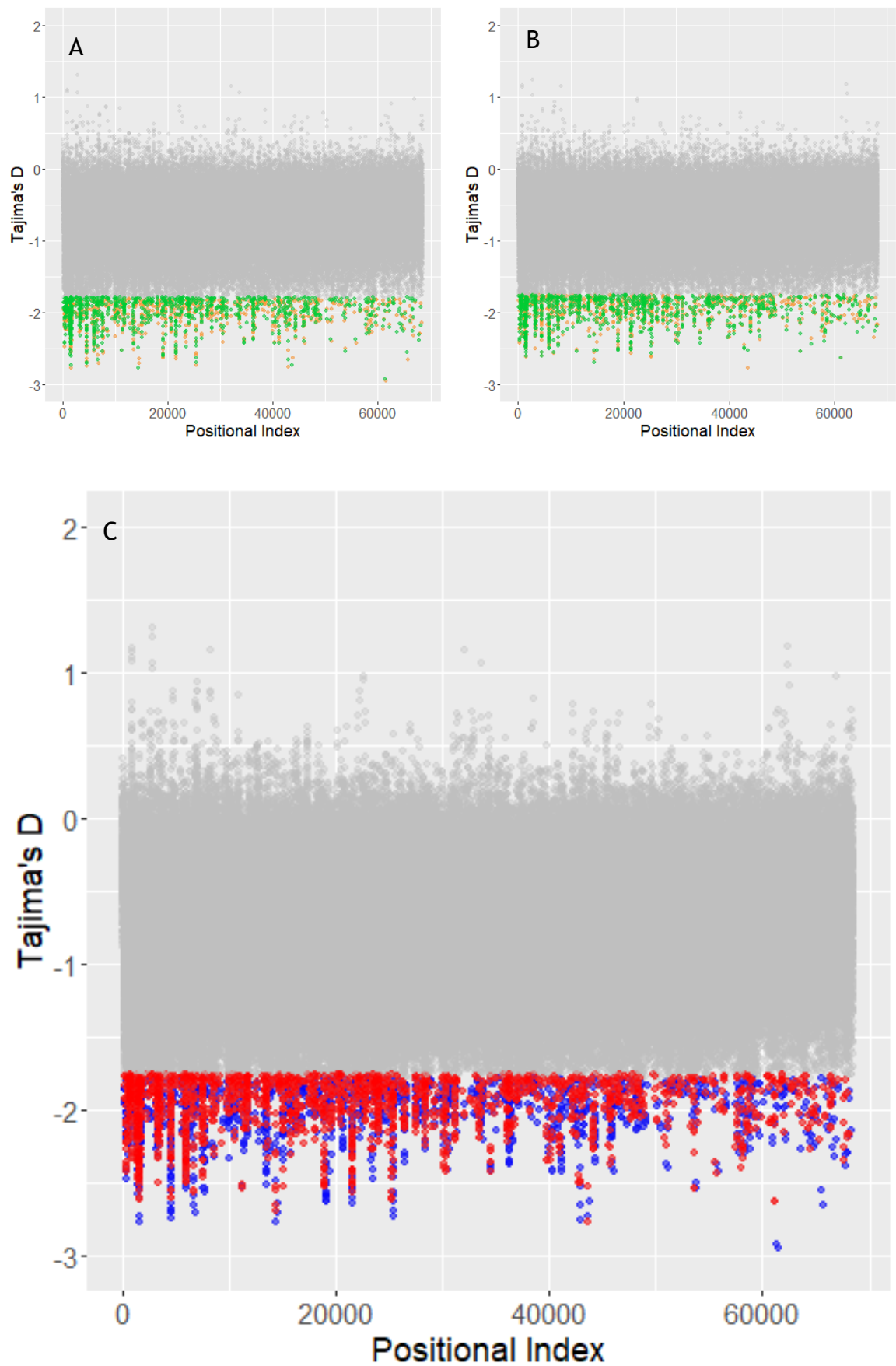


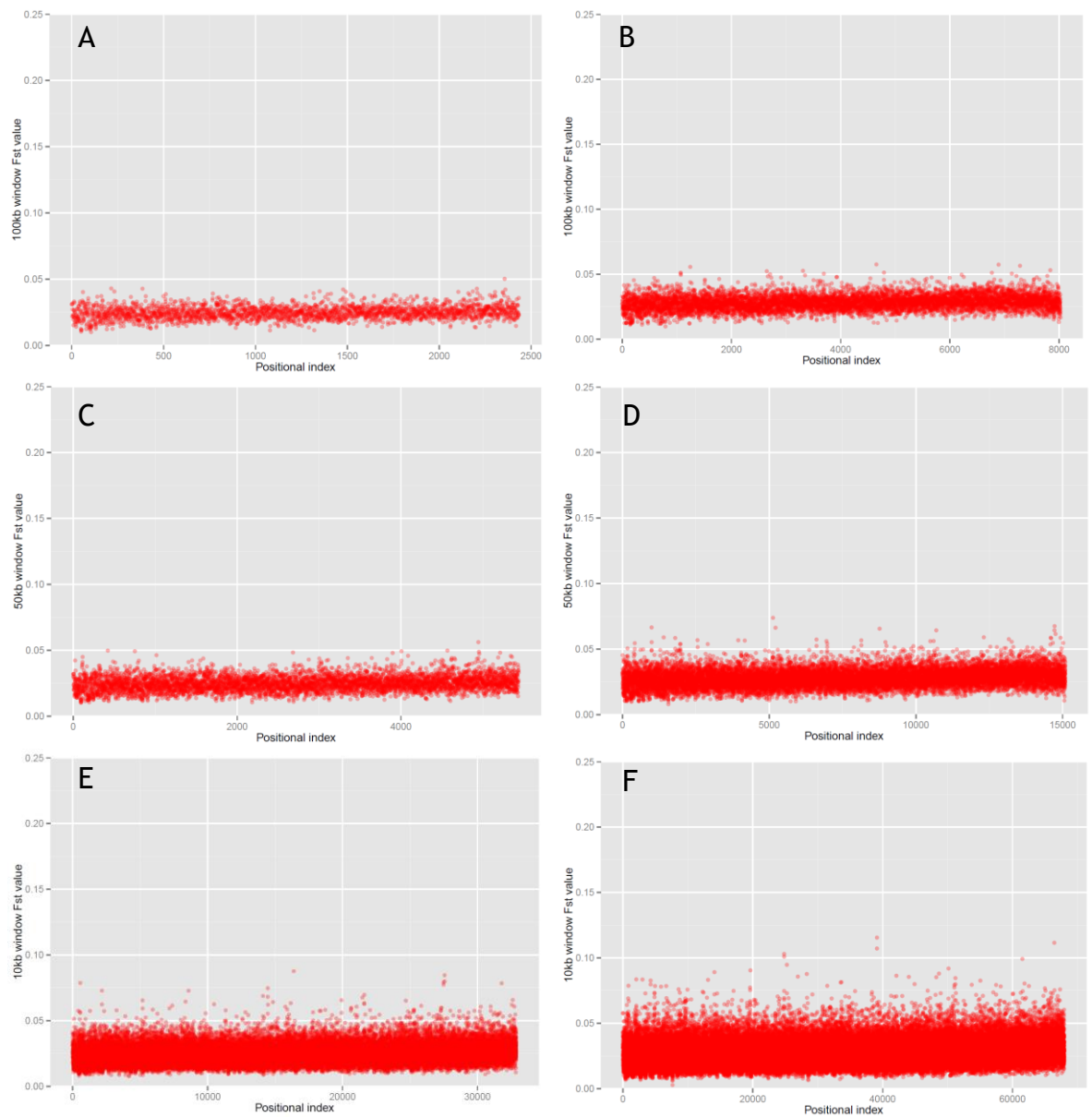
Figure 5.9: Farm 2 genome-wide Tajima's D. Each point represents one 10 kb window. The 1000 most negative windows are coloured. Key to colours: Pre-treatment = blue, Post-IVM = red.



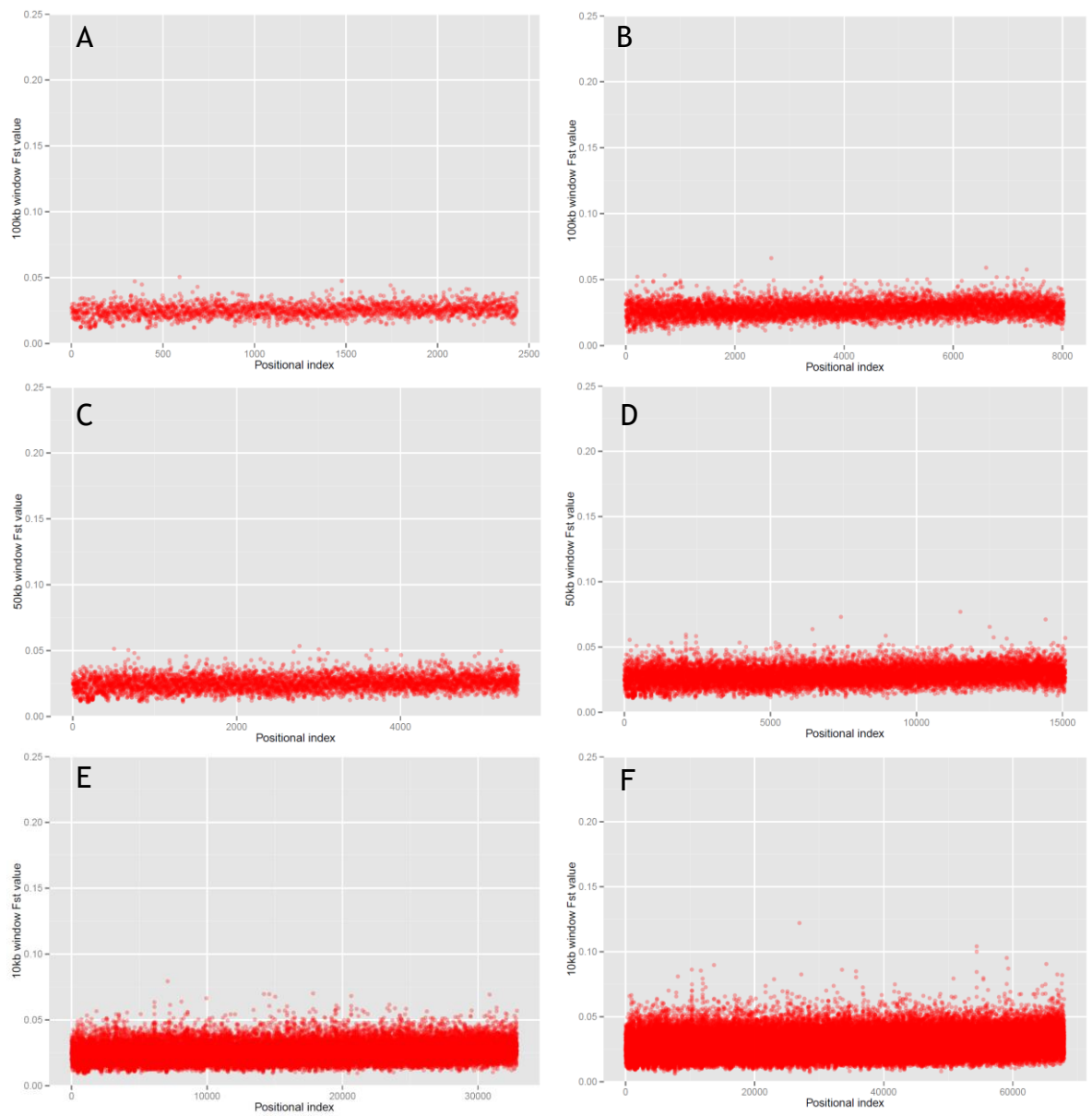
**Figure 5.10: Farm 3 genome-wide Tajima's D.** Each point represents one 10 kb window. The 1000 most negative windows are coloured. A. Farm 3 Pre-IVM replicates, B. Farm 3 Post-IVM replicates. For A and B, replicates are distinguished by colour. C. Farm 3 all samples (Pre-IVM = blue, Post-IVM = red).

### 5.3.5 Identification of differentiated genomic regions between pre- and post-treatment samples

Regions of genetic differentiation between the pre- and post-IVM treatment sample populations were identified using a fixation index measure:  $F_{ST}$  (Hartl and Clark, 2007; Kofler et al., 2011b). Parameters used to calculate pairwise  $F_{ST}$  values are detailed in the methods (section 2.9.4).  $F_{ST}$  is a measure of how fixed an allele is between samples within a population (Bird et al., 2011). As seen in Chapter 4, individual SNP  $F_{ST}$  values can often be quite high, but without any selection at the surrounding locus. The ddRAD-Seq data were too sparse to allow effective smoothing of the data, but Pool-Seq increased the coverage across the genome enabling confident smoothing of SNPs using distinct 10 kb windows to account for diversity and overinflated SNP  $F_{ST}$  values. Initially window sizes of 100 kb, 50 kb and 10 kb were tested (Figure 5.11 and Figure 5.12). As the window size reduced, more of the genome had potential to be included in the analysis (87.7 to 99.3%). Requiring complete coverage of a window returned no data, as all windows were excluded. The number of windows included in the analysis was greater for a window fractional coverage of 0.1 than 0.5, potentially due to variable coverage across a contig, either in absence of reads aligning, or in their depth. In addition, this reflects the number of contigs with total length less than 5 kb. The final window size of 10 kb, and the minimum window fractional coverage of 0.1 (i.e. 1000 bp), were chosen to maximise the number of contigs included in the analysis, whilst still permitting some smoothing of the data to account for diversity.



**Figure 5.11: Farm 2 Manhattan plots comparing different distinct window sizes and fractional coverages. All compare the pre-treatment sample with the post-ivermectin sample. A, C and E are 0.5 fractional coverage; B, D and F are 0.1 fractional coverage. A and B are 100 kb windows, C and D are 50 kb windows and E and F are 10 kb windows. Windows are plotted along the x-axis in order from the start of contig 1 to the end of the last contig for which values were calculated; contig 1 is the longest contig, and with some exceptions, contigs generally reduce in size from left to right.**

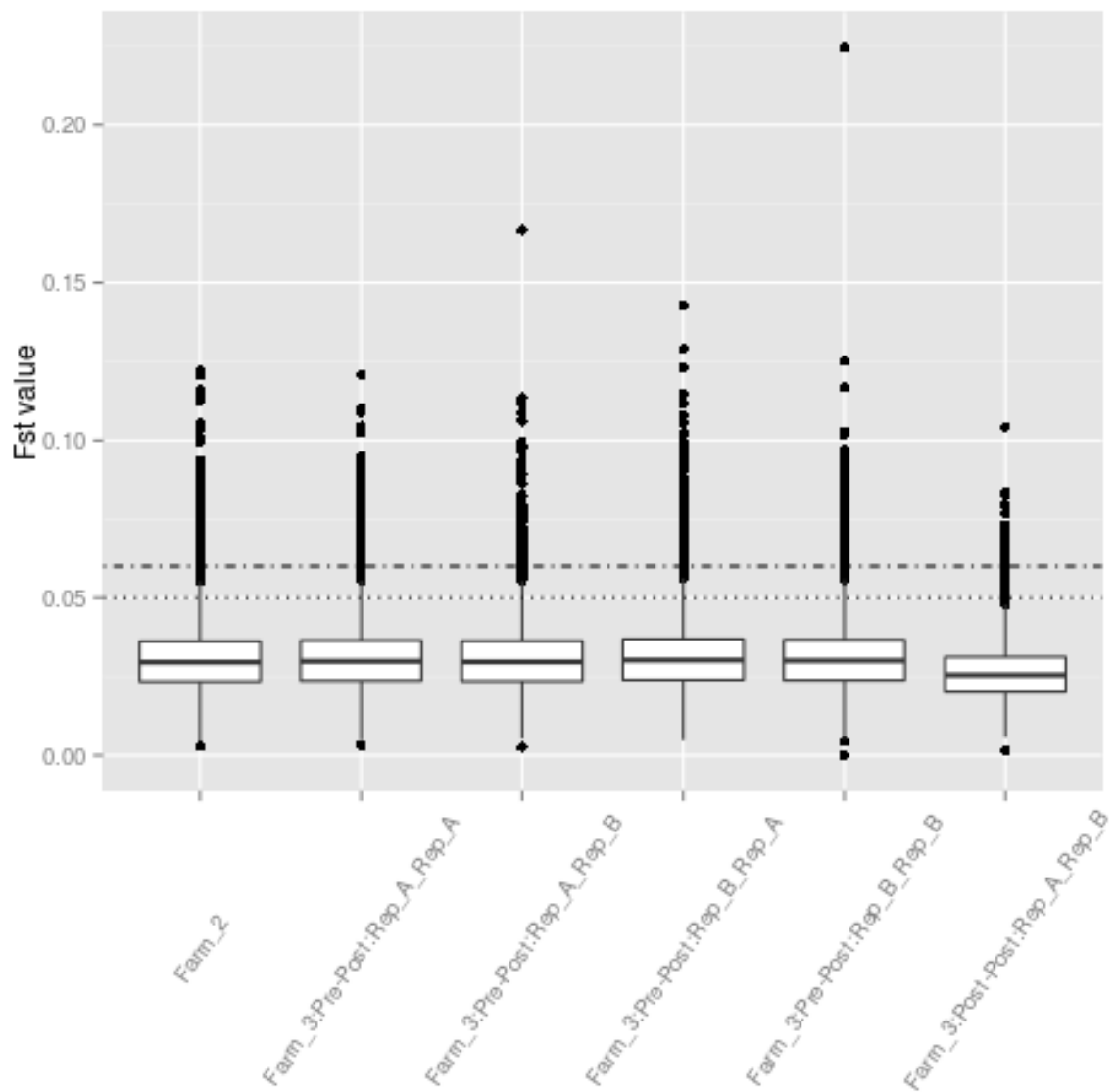


**Figure 5.12: Farm 3 Manhattan plots comparing different distinct window sizes and fractional coverages. All compare the pre-ivermectin sample with the post-ivermectin sample. A, C and E are 0.5 fractional coverage; B, D and F are 0.1 fractional coverage. The average  $F_{ST}$  value across all four replicate pre-post pairwise comparisons is plotted. A and B are 100 kb windows, C and D are 50 kb windows and E and F are 10 kb windows. Windows are plotted along the x-axis in order from the start of contig 1 to the end of the last contig for which values were calculated; contig 1 is the longest contig, and with some exceptions, contigs generally reduce in size from left to right.**

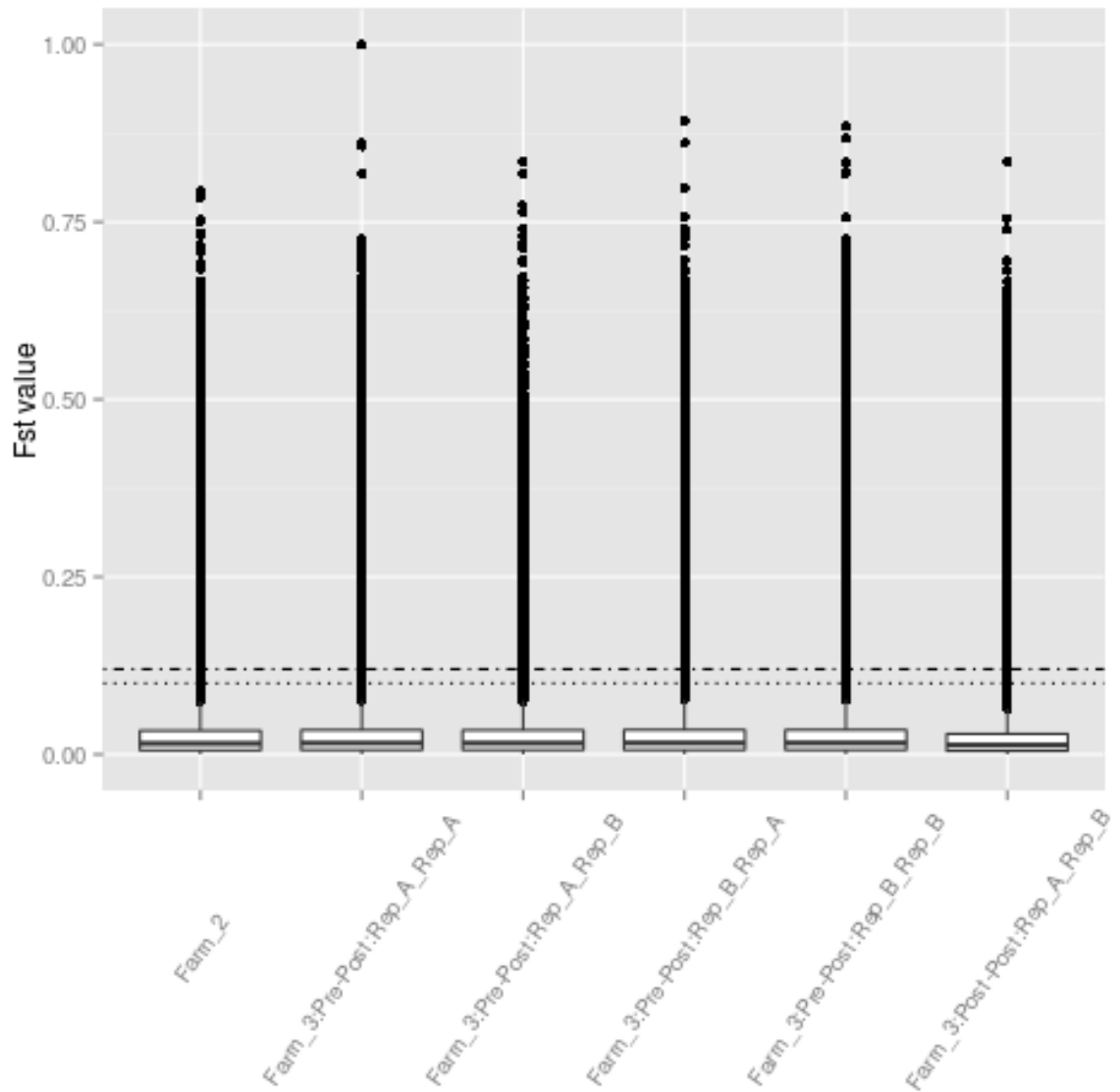
Following  $F_{ST}$  analysis of the full dataset, using the  $MAPQ \geq 20$  and  $BQ \geq 30$  filters, the results showed similar median  $F_{ST}$  values for each of the farm pre- to post-IVM pairwise comparisons, both using distinct 10 kb windows along the genome (Figure 5.13), and for individual SNP values (Figure 5.14). The mean and median values were lower for the SNPs than for the 10 kb window data. However, maximal values were considerably higher for SNPs than for the smoothed 10 kb windows. Values calculated for the Farm 3 post-IVM replicate A



vs post-IVM replicate B pairwise comparison were lower than for the pre- to post-IVM comparisons, as expected.



**Figure 5.13: 10 kb window  $F_{ST}$  pairwise comparisons. All pairwise comparisons performed are shown. The thresholds used for filtering differentiated regions are indicated. For all pre- to post-treatment comparisons (Farm 2 samples and Farm 3 replicate samples), the dot-dash line indicates the threshold above which windows were retained for further investigation. For each comparison this was very similar, and so just one line, at 0.06, is shown. The Farm 3:Post-Post technical replicate comparison had a lower threshold (0.05), indicated by the dotted line. Regions present as outliers in the Farm 3:Post-Post technical replicate comparison were excluded from further Farm 3 analyses as potential false positives.**



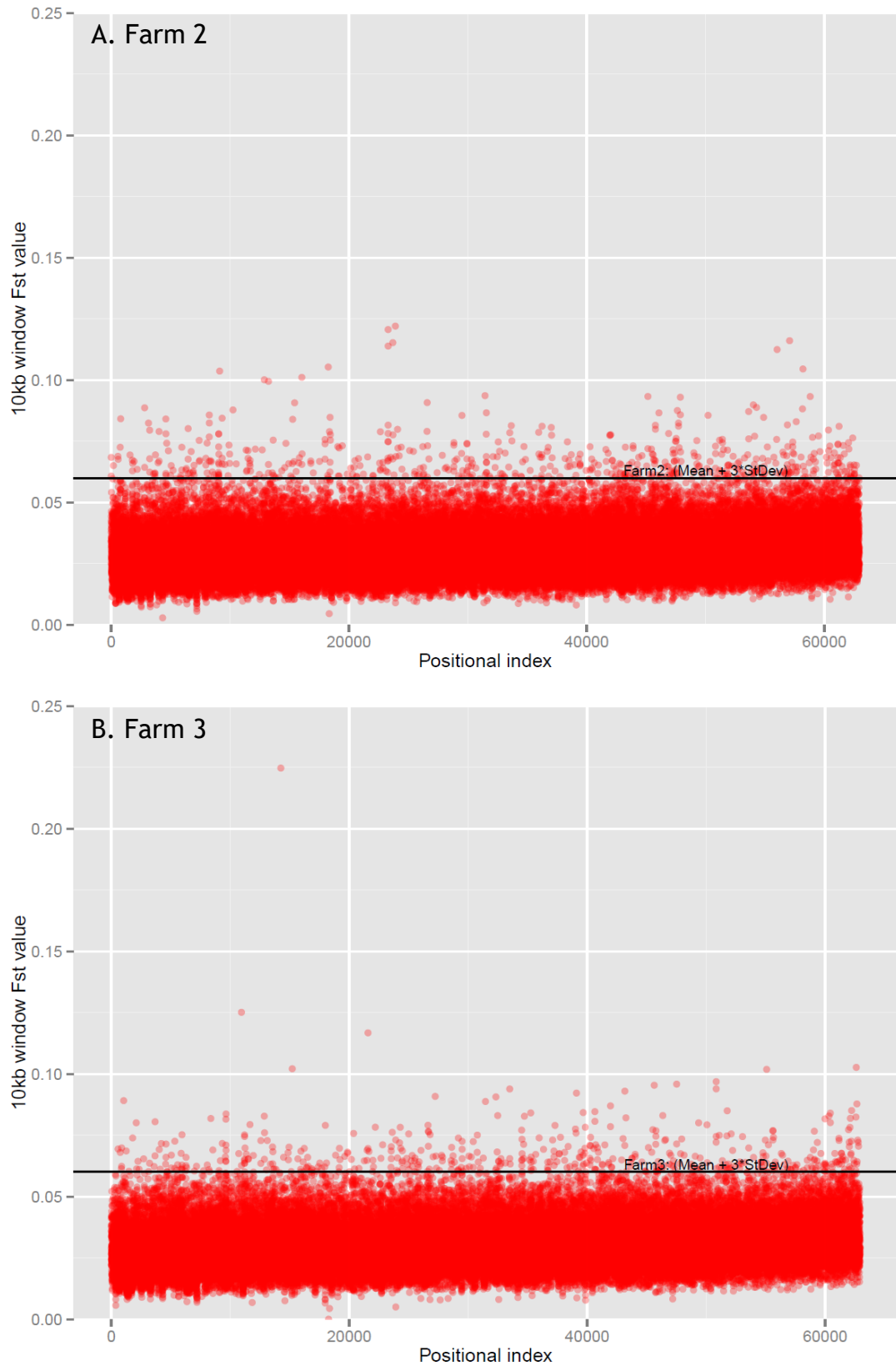
**Figure 5.14: SNP  $F_{ST}$  pairwise comparisons. All pairwise comparisons performed are shown. The thresholds used for filtering genetically differentiated SNPs are indicated. For all pre- to post-treatment comparisons, the dot-dash line indicates the threshold above which SNPs were retained for further investigation. For each comparison this was very similar, and so just one line, at 0.12, is shown. The Farm 3:Post-Post technical replicate comparison had a lower threshold (0.10), indicated by the dotted line. SNPs present as outliers in the Farm 3:Post-Post technical replicate comparison were excluded from further Farm 3 analyses as potential false positives.**

For each pairwise comparison a threshold was calculated, equivalent to an  $F_{ST}$  value of three standard deviations greater than the mean  $F_{ST}$  for that comparison (Table 5.5). SNPs and windows greater than the threshold were retained for further investigation. The presence of diversity within the population produces noise within the data and so for Farm 3, data were further filtered using the technical replicates as described in 2.9.4 (Robasky et al., 2014). Manhattan plots appeared similar for the pre- to post-IVM replicate comparisons on each farm (Figure 5.15). Outlier regions were distributed across the genome, and no single large locus under IVM selection was evident. There was no obvious difference in the 10 kb outlier  $F_{ST}$  values between farms (Figure 5.15).

**Table 5.5: Pool-Seq  $F_{ST}$  thresholds.**

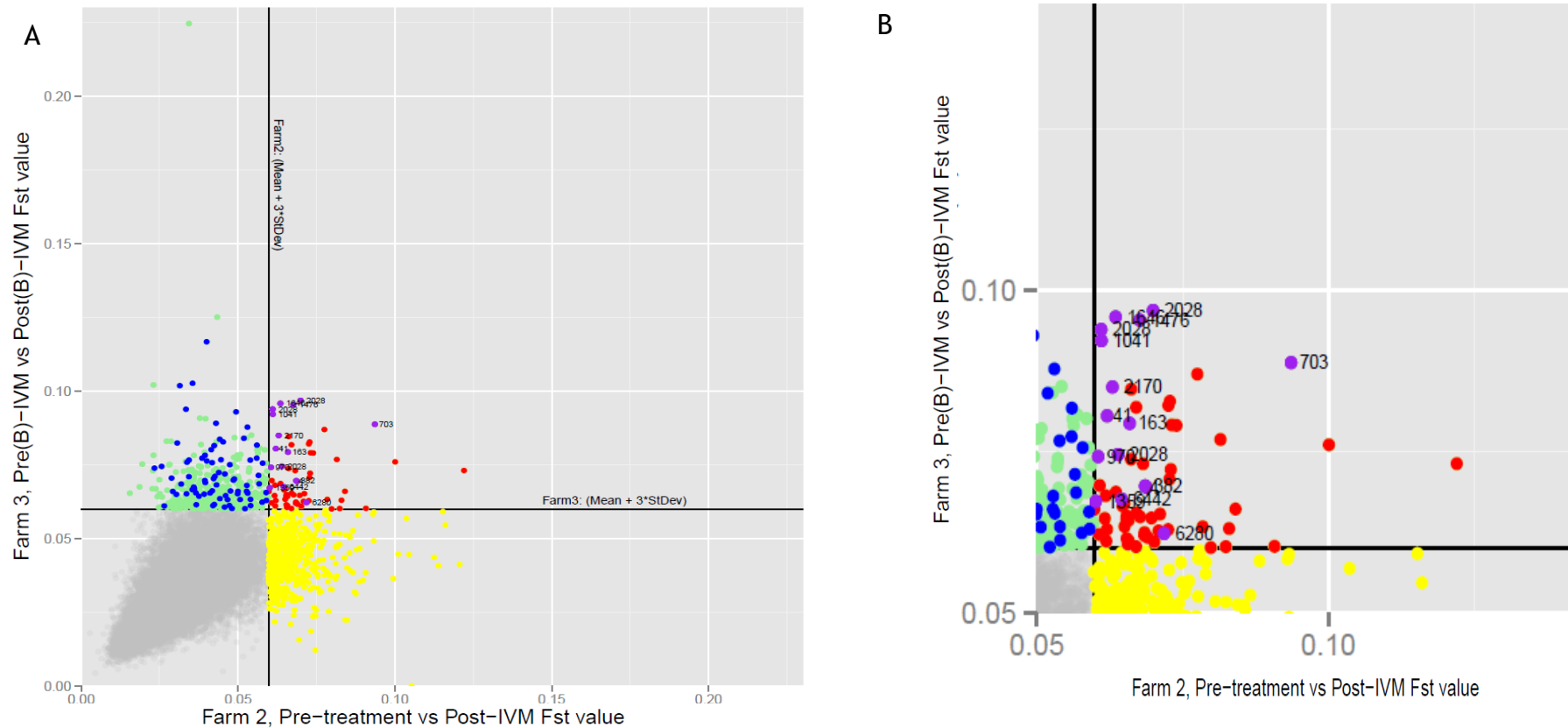
	10 kb window $F_{ST}$ Threshold	SNP $F_{ST}$ Threshold
Farm 2 pre-treatment vs post-IVM treatment	0.059770	0.116409
Farm 3 pre-IVM rep A vs post-IVM rep A	0.059739	0.117296
Farm 3 pre-IVM rep A vs post-IVM rep B	0.059374	0.116733
Farm 3 pre-IVM rep B vs post-IVM rep A	0.060321	0.118530
Farm 3 pre-IVM rep B vs post-IVM rep B	0.060064	0.118046
Farm 3 post-IVM rep A vs post-IVM rep B	0.050054	0.102495

For each pairwise comparison between pre- and post-treatment samples the threshold  $F_{ST}$  value, used to filter data to retain only the most highly differentiated regions and SNPs is shown. The threshold was calculated as three standard deviations above the mean  $F_{ST}$  value for that comparison.



**Figure 5.15: Manhattan plots of Farm 2 and Farm 3 Pool-Seq data A. Farm 2; pre-treatment compared with post-IVM sample. B. Farm 3; pre-IVM replicate B vs post-IVM replicate B pairwise comparisons are plotted. For each dataset the threshold used for filtering regions of interest is shown by a black line. Threshold is three standard deviations above the mean  $F_{ST}$  value.**

Regions of selection were difficult to visually determine using the Manhattan plots due, in part, to the large, fragmented and unordered genome assembly. Values calculated for each farm were plotted against each other to compare data between farms (Figure 5.16). When plotting each of the Farm 3 technical replicate comparisons against the Farm 2 data, it was evident that there were differences as to which windows were genetically differentiated on both farms. Also, differences were present between technical replicate comparisons. For clarity, just one technical replicate comparison is plotted in Figure 5.16, however points are coloured according to whether the  $F_{ST}$  value was considered high in all four technical replicate comparisons for Farm 3 or was only high in some. From this point forward, for Farm 3 analyses, only those windows considered of note in all four pre-post pairwise comparisons, and below the Farm 3:Post-Post technical replicate comparison threshold will be considered.



**Figure 5.16: Farm 2 vs Farm 3 10 kb window  $F_{ST}$  values. The Farm 3 pre-IVM replicate B vs post-IVM replicate B comparison is plotted. Each point is a 10 kb window. The threshold for each farm is indicated. Tci2 contig IDs are given. (B. is enlarged detail of A.) Key to colours: Purple = Window genetically differentiated on Farm 2 and in all Farm 3 replicate comparisons (but not in the Farm 3:Post-Post replicate comparison), Red = Window genetically differentiated on Farm 2 and in the plotted Farm 3 replicate comparison, Yellow = Window genetically differentiated on Farm 2 only, Blue = Window genetically differentiated in all Farm 3 replicate comparisons (but not in the Farm 3:Post-Post replicate comparison), Green = Window genetically differentiated in the plotted Farm 3 replicate comparison.**

In total, sixteen 10 kb windows had values greater than the  $F_{ST}$  thresholds for both Farm 2 and Farm 3 (Table 5.6, Appendix 5). Most of these windows were on separate contigs, however three were on contig 2028, spanning a region from 50 to 90 kbp (50-60 kbp, 70-90 kbp). The number of windows expected to be differentiated on both farms by chance was one ( $p < 0.001$ ).

Considering Farm 2 alone, a total of 576 10 kb windows were above the  $F_{ST}$  threshold, identified across 435 contigs (Table 5.6, Appendix 6). Several of these windows were adjacent to one another. Although no single large genetically differentiated region was apparent, contig 118 had 11 windows with raised  $F_{ST}$  values, with the longest consecutive region measuring 30 kb in length. The widest region overall was found on contig 422, measuring 40 kb, with a neighbouring differentiated region of 20 kb (separated by just 10 kb). All windows were ranked by decreasing  $F_{ST}$  value, and two of these 10 kb windows on contig 422 were within the top ten (Appendix 6).

A total of 91 10 kb windows were retained following filtering of Farm 3 data (Table 5.6, Appendix 7). These windows were distributed across 76 contigs. Again, several of these windows were adjacent to each other. However, in comparison to Farm 2, the maximum number of windows identified on a single contig was four (contig 129), and these were distributed such that the longest consecutive region was 20 kb in length, although a third window was identified 10 kb distant from this, suggesting that the overall region of selection might be 40 kb (Appendix 7).

The total number of SNPs retained with  $F_{ST}$  values higher than the threshold was considerably greater for Farm 2 (636,541) than Farm 3 (24,052, Table 5.6). Only 1901 SNPs were genetically differentiated in both farm datasets. A Fisher's Exact Test of the allele frequency difference between the pre- and post-treatment populations was performed using PoPoolation2 v1.201. Filtering SNPs to retain only those with a significant difference ( $p < 0.05$ ), retained over a million SNPs on Farm 2, but only 52,426 SNPs on Farm 3. When filtering SNPs using both  $F_{ST}$  values and allele frequency differences just 4.08% and 0.05% of all of those SNPs retained by  $F_{ST}$  alone were kept for Farm 2 and Farm 3 respectively. Excluding all SNPs not within a 10 kb window of high  $F_{ST}$  value further reduced this number (Table 5.6). It also changed the relative proportion of SNPs retained on each

farm, with 0.11% of SNPs of high  $F_{ST}$  value retained in the Farm 2 dataset, and 0.01% retained in the Farm 3 dataset. No SNPs were identified which had both a high  $F_{ST}$  value and a significant allele frequency difference in the sixteen 10 kb windows identified on both farms (Table 5.6). In addition to the  $F_{ST}$  analysis, a Cochran-Mantel-Haenszel test was performed (Kofler et al., 2011b; McDonald, 2014). Only a single SNP (on contig 331) was identified with a consistent allele frequency difference between both farms ( $p < 0.0001$ ). This SNP was not found within a 10 kb window of interest.

**Table 5.6: Counts of genomic regions and SNPs identified as being of interest on each of Farms 2 and 3.**

	Farm 2	Farm 3	Both farms
Total No. of 10 kb windows above the $F_{ST}$ threshold (Tci2 contigs)	576 (435)	91 (76)	16 (14)
Total No. of SNPs above the $F_{ST}$ threshold	636,541	24,052	1901
Total No. of SNPs with significant allele frequency difference	1,131,099	52,426	3298
Total No. of SNPs with both a high $F_{ST}$ and a significant allele frequency difference	26,002	11	1
Total No. of SNPs with both a high $F_{ST}$ and a significant allele frequency difference found within the high $F_{ST}$ 10 kb windows	676	2	0

Thresholds used were three standard deviations above the mean value of the pairwise comparison and are detailed in Table 5.5.

### 5.3.6 Comparison with ddRAD-Seq results

Contigs containing ddRAD-Seq SNPs identified in the previous experiment (Chapter 4), were compared to the contigs containing genetically differentiated 10 kb regions in the Pool-Seq experiment (Appendix 8 and 9). For both experiments, regions or SNPs compared were those with  $F_{ST}$  values of greater than three standard deviations above the mean  $F_{ST}$ . In total, 131 contigs were identified as potentially under selection by IVM on Farm 2 in both experiments. However, when comparing the ddRAD-Seq Farm 2 data with the Pool-Seq Farm 3 data, just 21 contigs appeared under selection. Just one RADlocus in the Farm 2 data was found within a Pool-Seq window. No RADloci were identified within Pool-Seq windows on Farm 3.



Of those manually annotated in 4.3.10, very few were also identified in the Pool-Seq experiment. On Farm 2 four contigs which had been annotated previously were identified by Pool-Seq. These were contigs 77, 86, 163 and 827.

Comparison of the relative distance between the RADlocus, and the Pool-Seq 10 kb window revealed that all four contigs had 80 kb or greater distance between the SNP and the Pool-Seq region. Thus, the manual annotation of the contigs, extending only approximately 75 kb distant from a RADlocus would not have identified genes within the Pool-Seq regions. Contig 163 was the only manually annotated contig identified in both experiments for Farm 3. The Pool-Seq window identified on contig 163 in the Farm 3 data was the same as in the Farm 2 data, and was approximately 80 kb distant from the RADlocus identified on Farm 2. However, the gene manually annotated at 75 kb distance from the RADlocus did not overlap with the Pool-Seq window. The Pool-Seq window contained a gene with sushi-domain, which may suggest involvement in extracellular recognition. The gene had an *H. contortus* homologue predicted by BLASTP but no *C. elegans* homologue.

## 5.4 Discussion

To identify markers of IVM resistance, Pool-Seq was performed using *T. circumcincta* progeny obtained pre- and post-IVM treatment as part of a FECRT. Two geographically separated farms, with differing reductions in *T. circumcincta* egg output post-IVM were selected. The genetic basis of IVM resistance in *H. contortus* appears similar across many different isolates (Doyle et al., 2019; Sallé et al., 2019), and the aim was to identify genetic markers which could be used diagnostically on many farms. For this reason, farm  $F_{ST}$  results were compared. However, many differences may exist between isolated populations unrelated to IVM selection (Sarai et al., 2014; Doyle et al., 2019; Sallé et al., 2019) and so comparisons to identify genetically differentiated regions of the genome, potentially under selection by IVM, were conducted within each farm population separately.

### 5.4.1 Whether a single locus, or multiple loci, contribute to IVM resistance on each farm is difficult to discern

Multiple 10kb windows appeared genetically differentiated between the *T. circumcincta* pre- and post-treatment populations on each farm. However, the relative genomic location of these windows, within the draft, fragmented Tci2 genome is not known. As such it is not possible to determine whether regions correspond to a single locus, or many loci. For example, Doyle et al. (2019) plotted their  $F_{ST}$  data (comparing a backcrossed, IVM-selected strain, with a susceptible isolate), on three separate versions of the *H. contortus* genome, assembled in various stages of completion from over 23,000 contigs (with an N50 of 83 kb and with 1,151 contigs longer than this length), to just eight scaffolds. As the genome assembly improved, it became possible to determine peaks of selection, whereas, with the same data, this was not possible on the original assembly.

$F_{ST}$  analysis indicated that only 160 kb, in total, was genetically differentiated on both farms, corresponding to regions on 14 different contigs. These results may indicate that the major mechanism of IVM resistance falls within those regions identified on both farms. Compared to the number of windows expected to be identified on both farms by chance ( $<1$ ), the observed number of windows (16) is significantly greater. This would add support to the theory of a single dominant locus conferring IVM resistance in adult worms of *T. circumcincta*. In *H. contortus*, a region on Chromosome V has been identified in multiple studies comparing IVM susceptible and resistant isolates (Sallé et al., 2019), or genetic crosses (Laing, *pers comm*, Redman et al., 2012; Doyle et al., 2019). This region is divided into two main peaks, spanning 5 Mb and 3 Mb. In contrast, 910 kb across 76 contigs appeared genetically differentiated on Farm 3 and 5.76 Mb were genetically differentiated on Farm 2, spread across more than 400 contigs. If IVM resistance in parasitic nematodes is polygenic, as is the case in *C. elegans* (Dent et al., 2000; James and Davey, 2009; Page, 2018), the genetic basis of IVM resistance may differ between farms, consistent with work on field populations of *Onchocerca volvulus* (Doyle et al., 2017). Genetically differentiated regions, which did not appear to be selected on both farms, often had stronger signals of selection than those that were common to both. This may indicate that the locus conferring IVM resistance lies within the higher  $F_{ST}$  regions - with a different

genetic basis of resistance on each farm. Alternatively, due to the dominant phenotype expected, the regions with lower  $F_{ST}$ , but common to both farms may be more important. This is because dominant phenotypes could allow maintenance of heterozygosity within the population, maintaining some diversity at the IVM resistance locus. It would be expected to be detected due to enrichment of alleles conferring resistance within the post-IVM population, but would not necessarily be the highest  $F_{ST}$  locus. Alternatively, multiple variants within the same gene may confer resistance to IVM, which would also maintain diversity - as was noted in BZ resistant *C. elegans* (Hahnel et al., 2018) and artemisinin resistant *Plasmodium falciparum* (Cheeseman et al., 2015).

In addition to the large region of selection on Chromosome V (Doyle et al., 2019; Sallé et al., 2019), other loci appear to be under IVM selection in *H. contortus* populations treated at half-doses of IVM (Laing, *pers comm*). On these farms it is entirely possible that multiple IVM resistance mechanisms have been selected over time. For example, from under-dosing of sheep (Sargison, 2011), or from selection of mechanisms which support multi-drug resistance and may be selected by multiple anthelmintics, such as the P-glycoproteins (Raza et al., 2016b). Furthermore, in the UK, with high levels of sheep movement between farms, each farm may have multiple origins of IVM resistance, increasing overall genomic diversity and producing soft selective sweep patterns, rather than harder sweeps. This has been observed for BZ resistance previously (Redman et al., 2015). This may be especially relevant on Farm 2, which is known to have recently merged two flocks.

As IVM has multiple targets (e.g. GluCl<sub>s</sub>, GABA receptors, FXR etc) and affects multiple different pathways in a diverse range of organisms (Holden-Dye and Walker, 1990; Jin et al., 2013; Ballesteros et al., 2016; Gallardo et al., 2018; Ménez et al., 2019), it is possible that more than one protective mutation may arise which confers some resistance to IVM in diverse field populations. In addition, different mechanisms of resistance may be important at different lifecycle stages - all of which may experience some selection pressure in field populations. Lastly, aside from genome wide diversity naturally occurring between populations, the presence of adaptive mutations, in the face of a necessary, but perhaps deleterious mutation conferring IVM resistance, may be

additionally selected for independently. Such mutations may overcome any fitness cost and increase genetic diversity (and apparent selection by IVM) within and between populations (Levin et al., 2000).

#### **5.4.2 Post-IVM samples are still highly diverse, but populations are less diverse than expected**

The genome-wide mean nucleotide diversity (Tajima's  $\pi$ ), was very similar between all samples and did not suggest considerable selection of a particular sub-population of individuals on either farm. In addition, allele frequency correlations (Pearson's correlation coefficient), suggest that all samples are equally diverse, without grouping of samples within farm or by treatment. Clearly individuals that are IVM resistant are as diverse within their sub-population as those which are susceptible. It is therefore unlikely that extensive LD over a region large enough to affect genome wide diversity measures has occurred by IVM selection. Alternatively, initial selection pressure on the population by IVM treatment may have occurred sufficiently distantly in time to allow subsequent outbreeding of any selected region within the population as a whole (both by recombination and by new mutations arising within the locus), supported by genome wide patterns of Tajima's  $D$ .

When comparing the Tajima's  $\pi$  estimates of *T. circumcincta* sample populations in this study with genome-wide estimates of several lab isolates of *H. contortus* (Doyle et al., 2019), *T. circumcincta* appears to be slightly more conserved than perhaps might be expected for such a highly diverse outbred population. Resistant strains of *H. contortus* (MHco4(WRS) and MHco10(CAVR)), and the susceptible inbred strain MHco3(ISE), had mean autosomal wide estimates of nucleotide diversity of between 0.022 and 0.038 (Doyle et al., 2019), comparable to or higher than the genome-wide mean estimates of *T. circumcincta* presented here (0.024). Nevertheless, the mean nucleotide diversity of the X chromosome of *H. contortus* was markedly lower, being at most 0.017 (MHco10(CAVR)), and as a minimum, only 0.008 (MHco3(ISE)) (Doyle et al., 2019). As the *T. circumcincta* genome is not assembled to chromosome level, it is not possible to tease out the differences relating to this disparity within the data, and the mean autosomal diversity may be higher than the genome-wide diversity calculated here. In addition, if some reads did not align

due to high diversity then this could artificially lower the genome wide Tajima's  $\pi$ .

### **5.4.3 Why might fewer reads have aligned post-treatment compared with pre-treatment?**

Pre-treatment populations on both farms had an increased number of sites with coverage depth greater than 60x compared with post-treatment populations, which had a greater proportion of sites with coverage depth less than 40x. This correlated with both a reduced percentage of reads aligning, and fewer reads aligned in total, in these post-IVM samples. Interestingly, a slightly smaller percentage of reads also aligned to the genome for the Farm 2 post-treatment ddRAD-Seq experiment compared with the pre-treatment sample (4.3.3). This may reflect an increased genetic difference between the post-IVM populations and the MTci2 population used to assemble the reference genome. MTci2 was isolated from the field before IVM was licensed and is fully susceptible to a standard dose of IVM (Skuce et al., 2010). Although the sample size is small, it is interesting to speculate if this phenomenon is related to genetic differentiation (Kofler et al., 2016b), rather than merely stochastic chance, as it could indicate a loss of data, perhaps one or two alleles, specifically relating to IVM resistance. Although the extent of such an effect depends on how much of the genome is affected by IVM selection, a loss of diversity in the post-treatment populations due to a reduction in reads aligning, would still be likely to produce a signal.

In addition to the lower percentage of reads aligning in these post-treatment samples, the mean and outlier Tajima's  $D$  values were always slightly less negative, indicating fewer rare alleles in the post-treatment samples compared with the pre-treatment. This might imply selection upon these populations by IVM. Alternatively, this effect may have been caused by certain alleles not aligning to the MTci2 genome in these samples.

### **5.4.4 Sample coverage was lower than expected – and may have affected sample size, increasing differences between technical replicates**

In order to increase the chance of truly sampling allelic diversity within the population, and obtain accurate allele frequencies, we aimed to use Pool-Seq

with as many individuals as possible, while keeping the numbers of individuals used for each pool consistent within and between farms. The cost-benefit of Pool-Seq increases as the number of individuals included in the pool increases (Gautier et al., 2013a) and variation in DNA input quantity between individuals is reduced (Rode et al., 2018). In this study, we were able to use three times as many worms for each sampling time point as for the ddRAD-Seq experiment. Ninety-one individuals were used per pool and a coverage of ~100x aimed for (Rode et al., 2018), based on the expected genome size. This level of coverage would have allowed each individual to be sequenced once, providing a coverage per chromosome ( $2n$  = diploid) of 0.5x. In reality, between 70.4 to 75.4x coverage of each pool was obtained. However, following alignment of reads to the genome, actual coverage was much lower, ranging from 45.7 to 51.5x genome wide, with variability in coverage across the genome such that the mean mapped depth per site was between 35.5 and 39.2x (Table 5.4). This raises two concerns. Firstly, due to the actual pool coverage (0.27x per chromosome), as opposed to the expected coverage (0.5x), the sample size is much smaller than hoped. In a highly diverse population this could reduce the ability to accurately determine allele frequencies and hence detect selection. Secondly, it could have increased differences between technical replicates. These were included for Farm 3 and proved useful in reducing noise generated by false positive signals unrelated to IVM selection (Robasky et al., 2014). Although the Pearson's correlation coefficient was higher between technical replicate samples than between different biological samples, it was still only about 0.73. By reducing the sample size sequenced, the chance of stochastically sampling chromosomes within the pool, and limiting selection of allele frequencies within a sample, could increase differences between technical replicates. That the strength of IVM selection was strong enough to overcome this potential variation was evidenced by the reduced median  $F_{ST}$  values identified, and threshold calculated, for the Farm 3:Post-Post technical replicate comparison compared with the Pre-Post comparisons.

Reasons for the reduced relative to the expected coverage include contamination within the sample, suggested by the Kraken report and the subset of high GC reads within the Farm 3 post-IVM samples (Barbu et al., 1956). As mentioned above, the potential for diversity between MTci2 and the field

populations could lead to reduced numbers of reads aligning - only between 60 to 70% of reads aligned overall for each sample. Conversely, it is possible that part of the reason for these reads not aligning is that sections of the *T. circumcincta* genome could be missing from the Tci2 assembly. Lastly, of those reads which aligned, almost one fifth had a mapping quality of zero, indicating they aligned equally well in more than one position. This may indicate the presence of haplotypic regions within the draft genome, spreading reads over multiple contigs, which should instead be aligned in just one region. This would both reduce mean coverage across the genome and could also affect  $F_{ST}$  analysis.

## 5.5 Conclusion

This study has validated a Pool-Seq method to detect regions of differentiation across the genome in outbred UK field populations of *T. circumcincta*. Despite using very low quantities of input DNA, sequencing was successful, and quality was good. Coverage could have been higher, which has limited sample size. However, due to the draft nature of the genome, artefactual haplotypic diversity may still be present within the genome. Dependent on the extent of this, genome-wide sequencing coverage may be higher than that observed, and could therefore increase as the genome assembly improves. High nucleotide diversity, similar to that of *H. contortus*, was observed. Although slightly lower than perhaps expected, values calculated support the concept of a small genomic footprint of selection by IVM, with maintenance of a diverse population. No single strong signals of selection were detected, although these may have been masked by the fragmented state of the genome. Nevertheless, several regions appeared differentiated on both farms, suggesting that a common genetic basis of IVM resistance may reside within them. Alternatively, many regions of genetic differentiation were identified on each farm individually, which may support a multigenic basis of IVM resistance. All regions require further analysis as it is not possible to determine those which are selected by IVM from those selected by genetic hitch-hiking, or false positive signals of selection.

## 6 Sifting the wheat from the chaff: Identifying genes which might be associated with IVM resistance

### 6.1 Abstract

Genome wide association studies can identify many regions that are differentiated, without always providing a clear indication of which regions are the most clinically relevant. Using results obtained following Pool-Seq of pre- and post-treatment *Teladorsagia circumcincta* offspring collected from two UK farms, I performed further bioinformatic analyses to investigate regions potentially under IVM selection. I annotated the draft Tci2 *T. circumcincta* reference genome using MTci2 RNA-Seq data provided by the BUG Consortium. The lipophilic drug IVM induces a flaccid somatic paralysis and can induce starvation and inhibit development associated with pharyngeal paralysis. In addition, IVM suppresses egg output of sensitive strongyle nematodes. I identified multiple regions containing neuronal genes, including one glutamate gated chloride channel, *glc-3* - although this was only identified on one farm. Some genes were predicted to be specific to males or females within highly differentiated regions. Multiple metabolic genes were identified, including those involved with lipid metabolism. In addition to within-farm analysis, regions identified as under selection were compared between farms. When separate Tci2 annotated genes were differentiated on each farm but the same homologue was predicted for both Tci2 genes, these genes were noted. My results were compared with other published GWAS of IVM resistance to identify whether genes were commonly selected across species. Lastly, copy number variation was assessed to specifically investigate P-glycoproteins, which have previously been indicated in anthelmintic resistance as a multi-drug resistant mechanism. Although many genes identified within regions of differentiation could potentially be associated with IVM resistance, no candidate genes were strongly linked with IVM treatment. In addition, evidence of multiple gene copies, representing technical artefacts within the available reference genomes was apparent. Results should therefore be carefully interpreted as the extent to which this artefactual duplication, and the fragmented state of the reference genome might impact upon this analysis is unknown.



## 6.2 Introduction

In the previous chapters, genetically differentiated regions were identified between progeny of pre- and post-IVM *T. circumcincta* populations on two separate farms. These regions were spread throughout the fragmented Tci2 genome and it was not possible to determine whether one or many loci were differentiated (Doyle et al., 2019). There were too many regions for detailed functional validation using molecular techniques. This chapter describes work investigating these genomic regions using bioinformatic analyses, highlighting regions that may be suitable for further study, based on the current literature.

Once genes of interest are identified, there are now many tools available for comparative genomics. An excellent database (WormBase) containing genes, proteins, phenotypes and gene interaction information for the closely related model nematode *C. elegans* is available (Howe et al., 2016). This can be used with WormBase ParaSite (Howe et al., 2017) to identify putative homologues of parasitic nematode genes and their potential function. In addition, WormBase ParaSite provides genomic and transcriptomic data for many related parasites, including the newly completed chromosomal assembly of *H. contortus*, PRJEB506 (MHco3ISE\_4.0). A published *T. circumcincta* genome, PRJNA72569 (T\_circumcincta.14.0.ec.cg.pg), is also available on WormBase ParaSite, which, although highly fragmented (81,734 scaffolds), has gene annotations and some protein information (Choi et al., 2017).

In January 2019, MTci2 RNA-Seq data generated by the BUG Consortium was used to annotate the Tci2 draft genome. Using WormBase, WormBase ParaSite and the literature, annotated genes within genetically differentiated regions were explored for potential relevance to IVM resistance. In addition, other recently published GWAS investigating IVM resistance in *T. circumcincta* (Choi et al., 2017), *H. contortus* (Sallé et al. (2019) and unpublished work by the BUG Consortium), and *Onchocerca volvulus* (Doyle et al., 2017) were compared with the present study. The aim was to identify consistent patterns, and genes or gene families, which may be important in understanding IVM resistance.

## 6.3 Results

### 6.3.1 Preliminary annotation of the Tci2 genome

MTci2 RNA-Seq data, of good quality, was provided by the BUG Consortium. This was used to guide annotation of the Tci2 genome as part of this PhD project. Approximately 87 to 90% of the reads aligned successfully and, in total, 55,014 putative genes were annotated by BRAKER v2.0 (Stanke et al., 2006; Stanke et al., 2008; Lomsadze et al., 2014; Hoff et al., 2016), with 60,450 transcripts in total. In contrast, the published *T. circumcincta* PRJNA72569 genome has 25,572 coding genes annotated (Choi et al., 2017), while the *H. contortus* chromosomal assembly, PRJEB506, (Howe et al., 2017) has 19,430 coding sequences and 20,707 gene transcripts. *C. elegans* (PRJNA13758, Howe et al., 2016) has 20,213 coding genes annotated, with a further 25,050 non-coding genes and over 61,000 gene transcripts in total. Many of the Tci2 annotated genes were fragmented or were missing start or stop codons. These results are as expected with a fragmented and haplotypic draft assembly within which genes may be split over multiple contigs and multiple alleles of the same gene may be present. It is therefore expected that multiple Tci2 annotated genes may be identified as homologues of the same gene in *H. contortus* or *C. elegans* or to be multiple copies of the equivalent gene in PRJNA72569. The published *T. circumcincta* genome is itself highly fragmented with many genes incompletely annotated, and some genes are known to be missing from the assembly.

### 6.3.2 Individual farm analysis identifies many genes of potential interest in IVM resistance

Regions of genetic differentiation were identified using 10kb distinct windows along the genome, allowing for potentially 68.5 million windows to be analysed. In total just 576 windows were above the  $F_{ST}$  threshold on Farm 2, and just 91 on Farm 3. Fewer were identified in the Farm 3 data due to an increased stringency of filtering using technical replicates (see section 2.9.4 for more detail). Genes were present within 438 of the 576 10kb windows above the  $F_{ST}$  threshold on Farm 2 ('genetically differentiated') and within 68 of the 91 genetically differentiated windows on Farm 3 (Table 6.1). A BLASTP search (Altschul et al., 1990) with the translated nucleotide sequences was used to identify homologues

in *C. elegans* and *H. contortus* and equivalent proteins (i.e. the same protein or a closely related paralogue) in PRJNA72569 as described in 2.9.5.

In total, 603 Tci2 genes were identified within these windows in the Farm 2 data, and for 526 of these, homologous or equivalent proteins were identified in at least one database (Table 6.1). Just over a third of these appeared to be parasite specific, with no homologue identified in *C. elegans*. Just over 2% of proteins were solely identified in the *T. circumcincta* PRJNA72569 database.

In total, 85 Tci2 genes were identified on Farm 3, and 73 of these equated to proteins identified in at least one database. A greater proportion of these were parasite specific (42.3%) compared with Farm 2, however none were solely identified in the *T. circumcincta* PRJNA72569 database.

**Table 6.1: Numbers of Tci2 annotated genes and their predicted proteins within genetically differentiated regions.**

	Farm 2	Farm 3
10 kb windows above the $F_{ST}$ threshold (Tci2 contigs)	576 (435)	91 (76)
Tci2 genes identified within these windows (transcripts)	603 (690)	85 (89)
Tci2 genes with at least one homologous or equivalent protein predicted	526 (87.2%)	73 (80.2%)
<i>Teladorsagia circumcincta</i> PRJNA72569 reference genome proteins predicted (% of all Tci2 genes)	324 (57.5%)	42 (50.6%)
<i>Haemonchus contortus</i> proteins predicted (% of all Tci2 genes)	455 (84.1%)	69 (81.2%)
<i>Caenorhabditis elegans</i> proteins predicted (% of all Tci2 genes)	283 (52.7%)	35 (43.5%)
Tci2 genes for which proteins were predicted in all three datasets (% of all Tci2 genes)	255 (42.3%)	28 (32.9%)
Tci2 genes for which parasite proteins were predicted, without a predicted homologue in <i>Caenorhabditis elegans</i> (% of all Tci2 genes); putatively parasite-specific genes	212 (35.2%)	36 (42.3%)
Tci2 genes for which proteins were identified solely in the PRJNA72569 database (% of all Tci2 genes); putatively <i>Teladorsagia circumcincta</i> -specific genes	12 (2.0%)	0 (0.0%)
Tci2 genes for which no protein was identified (% of all Tci2 genes)	77 (12.8%)	12 (14.1%)
10 kb windows with no Tci2 gene (% of all windows)	138 (24.0%)	23 (25.3%)

To ascertain protein function, gene ontology (GO) terms were extracted for the 526 proteins identified in the Farm 2 analysis, and the 73 proteins identified in the Farm 3 analysis, as described in 2.9.6. It was assumed that the GO term annotations would be most complete in the *C. elegans* genome assembly. For parasite-specific genes, GO terms were preferentially identified in *H. contortus* homologues rather than PRJNA72569 equivalent genes, if available.

Following extraction of GO terms and gene descriptions, search terms were used to identify genes of interest in the following categories; neuronal (42 genes in the Farm 2 data (Table 6.2) and three genes in the Farm 3 data (Table 6.3)), lipid metabolism (39 genes in the Farm 2 data (Table 6.4) and seven genes in the

Farm 3 data (Table 6.5)), female (13 genes in the Farm 2 data (Table 6.6) and three genes in the Farm 3 data (Table 6.7)) and male (seven genes in the Farm 2 data (Table 6.8) and just a single gene, *snf-10*, in the Farm 3 data (Table 6.9)). These categories were chosen because of the known effects of IVM on locomotion (Demeler et al., 2010a), feeding (Alvarez-Sanchez et al., 2005b) and reproduction (Sutherland et al., 2003a). The category of lipid metabolism was chosen as IVM is lipophilic (Sutherland and Campbell, 1990) and can affect lipid storage (Smus et al., 2017). It is also possible that due to starvation induced by IVM, altered lipid metabolism and storage of lipids may be important in survival in the presence of IVM (Laing et al., 2012). Fewer genes were identified in these categories in Farm 3 data than in Farm 2 data, reflecting the smaller number of genes overall. The complete GO term results for all genes identified within regions of genetic differentiation are presented in Appendices 10 and 11.

**Table 6.2: Farm 2 predicted homologues annotated with GO terms associated with neuronal tissues or processes.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00000054	<i>acr-15</i>	Acetylcholine Receptor
WBGene00000056	<i>acr-17</i>	Acetylcholine Receptor
WBGene00022104	<i>acsd-1</i>	2-amino-3-carboxymuconate-6-semialdehyde decarboxylase
WBGene00011258	<i>best-15</i>	Bestrophin homolog 15
WBGene00206487	<i>best-19</i>	Bestrophin homolog
WBGene00012779	<i>cdkl-1</i>	Cyclin-dependent kinase-like 1
WBGene00000528	<i>clh-1</i>	Chloride channel protein
WBGene00014054	<i>dbt-1</i>	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial
WBGene00000951	<i>deg-3</i>	Acetylcholine receptor subunit alpha-type <i>deg-3</i>
WBGene00000955	<i>des-2</i>	Acetylcholine receptor subunit alpha-type <i>des-2</i>
WBGene00001519	<i>gar-3</i>	Muscarinic acetylcholine receptor <i>gar-3</i>
WBGene00001593	<i>glc-3</i>	Glutamate-gated Chloride channel; Glutamate-gated chloride channel subunit
WBGene00001664	<i>gpa-2</i>	Guanine nucleotide-binding protein alpha-2 subunit
WBGene00001665	<i>gpa-3</i>	Guanine nucleotide-binding protein alpha-3 subunit
WBGene00010593	<i>gsnl-1</i>	Gelsolin-like protein 1
WBGene00006504	<i>kcc-1</i>	K <sup>+</sup> /Cl <sup>-</sup> -Cotransporter
WBGene00002234	<i>kqt-2</i>	Potassium channel, KvQLT family
WBGene00022283	<i>lgc-27</i>	Ligand-Gated ion Channel
WBGene00017314	<i>lgc-39</i>	Ligand-Gated ion Channel
WBGene00008022	<i>lgc-41</i>	Ligand-Gated ion Channel
WBGene00021437	<i>lgc-42</i>	Ligand-Gated ion Channel
WBGene00003029	<i>lin-44</i>	Abnormal cell lineage protein 44
WBGene00003058	<i>lov-1</i>	Location of vulva defective 1
WBGene00015478	<i>mapk-15</i>	Mitogen-activated protein kinase 15
WBGene00003404	<i>mpz-1</i>	Multiple PDZ domain protein

Gene stable ID	Gene name	Gene description or GO term
WBGene00017437	<i>nmgp-1</i>	Neuronal Membrane Glycoprotein
WBGene00003909	<i>pag-3</i>	Zinc finger protein PAG-3
WBGene00003878	<i>pept-3</i>	Peptide transporter 3
WBGene00004032	<i>pkc-1</i>	Protein kinase C-like 1B
WBGene00004212	<i>ptl-1</i>	Microtubule-associated protein
WBGene00019891	<i>R05F9.7</i>	Sensory perception of chemical stimulus
WBGene00006476	<i>rhgf-2</i>	RHo Guanine nucleotide exchange Factor
WBGene00008183	<i>rin-1</i>	RIN (Ras/Rab Interactor) homolog
WBGene00011392	<i>sbt-1</i>	Seven B Two (Mammalian 7BT prohormone convertase chaperone) homolog
WBGene00004902	<i>snf-3</i>	Sodium- and chloride-dependent betaine transporter
WBGene00004926	<i>snt-6</i>	Synaptotagmin
WBGene00005713	<i>srv-2</i>	Serpentine receptor class gamma
WBGene00011326	<i>T01D3.1</i>	Motor neuron axon guidance
WBGene00006615	<i>trp-2</i>	TRP (Transient receptor potential) channel family; TRP homologous cation channel protein
WBGene00006768	<i>unc-32</i>	Probable V-type proton ATPase 116 kDa subunit a
WBGene00021305	<i>zig-11</i>	2 (Zwei) IG domain protein
<i>HCON_00090710</i>	NA	Neurotransmitter:sodium symporter activity

All genes were identified using GO terms returned by the BIOMART tool (Smedley et al., 2015) in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART. Those for which no *C. elegans* homologue could be found were identified by *H. contortus* GO term annotations.

**Table 6.3: Farm 3 predicted homologues annotated with GO terms associated with neuronal tissues or processes.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00002976	<i>lev-9</i>	Synapse related
WBGene00003878	<i>pept-3</i>	Peptide transporter 3
WBGene00004909	<i>snf-10</i>	Transporter

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART.



Table 6.4: Farm 2 predicted homologues annotated with GO terms associated with lipid metabolism or storage.

Gene stable ID	Gene name	Gene description or GO term
WBGene00000020	<i>abt-2</i>	ABC Transporter family
WBGene00007112	<i>B0035.13</i>	Lipid metabolic process
WBGene00015544	<i>C06E8.5</i>	Lipid binding
WBGene00008291	<i>C54C8.3</i>	Hexosyltransferase
WBGene00000509	<i>cka-1</i>	Choline Kinase A
WBGene00000837	<i>cul-2</i>	Cullin-2
WBGene00000838	<i>cul-3</i>	Cullin-3
WBGene00014054	<i>dbt-1</i>	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial
WBGene00006483	<i>dgk-4</i>	Diacylglycerol kinase
WBGene00017373	<i>F10G7.10</i>	Ubiquitin protein ligase activity
WBGene00009237	<i>F28H7.3</i>	Lipid metabolic process
WBGene00018175	<i>F38B6.6</i>	Protein O-mannosyl-transferase TMTC1
WBGene00019029	<i>F58B6.1</i>	Lipid metabolic process
WBGene00011321	<i>fil-1</i>	Fasting Induced Lipase
WBGene00001511	<i>fzy-1</i>	WD repeat-containing protein <i>fzy-1</i>
WBGene00019612	<i>gpcp-1</i>	Putative glycerophosphocholine phosphodiesterase GPCPD1 homolog 1
WBGene00003478	<i>mtm-6</i>	Myotubularin-related protein 6
WBGene00020679	<i>ogdh-1</i>	2-oxoglutarate dehydrogenase, mitochondrial
WBGene00003877	<i>pept-1</i>	Peptide transporter family 1
WBGene00004032	<i>pkc-1</i>	Protein kinase C-like 1B
WBGene00004076	<i>pod-2</i>	Fatty acid biosynthetic process
WBGene00004205	<i>psr-1</i>	Bifunctional arginine demethylase and lysyl-hydroxylase <i>psr-1</i>
WBGene00004259	<i>pyr-1</i>	CAD protein Glutamine-dependent carbamoyl-phosphate synthase Aspartate carbamoyltransferase dihydroorotase
WBGene00011040	<i>R05H5.5</i>	Sphingolipid metabolic process

Gene stable ID	Gene name	Gene description or GO term
WBGene00011935	<i>scrm-1</i>	Phospholipid scramblase
WBGene00004926	<i>snt-6</i>	Synaptotagmin
WBGene00020393	<i>T10B5.7</i>	Lipid metabolic process
WBGene00020784	<i>tat-4</i>	Phospholipid-transporting ATPase
WBGene00006739	<i>ulp-4</i>	Ubiquitin-like protease 4
WBGene00020962	<i>W02H5.8</i>	Glycerone kinase activity
WBGene00022532	<i>ZC155.4</i>	Glycerophosphodiester phosphodiesterase activity
<i>HCON_00053760</i>	NA	Ubiquitin protein ligase activity
<i>HCON_00130960</i>	NA	Lipase
<i>HCON_00156530</i>	NA	Glycerophosphodiester phosphodiesterase activity
<i>TELCIR_04843</i>	NA	Triacylglycerol lipase
<i>TELCIR_05535</i>	NA	DAK2 domain protein
<i>TELCIR_13687</i>	NA	Diacylglycerol kinase accessory domain protein
<i>TELCIR_16739</i>	NA	Dehydrogenase E1 component
<i>TELCIR_21800</i>	NA	Oxoglutarate dehydrogenase (succinyl-transferring) activity

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART. Those for which no *C. elegans* homologue could be found were identified by *H. contortus* GO term annotations, or if they had only an equivalent gene identified in PRJNA72569, they were identified by the PRJNA72569 GO term annotations.

**Table 6.5: Farm 3 predicted homologues annotated with GO terms associated with lipid metabolism or storage.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00012019	<i>dkf-2</i>	Serine/threonine-protein kinase <i>dkf-2</i>
WBGene00017084	<i>E01A2.1</i>	Glutamate-cysteine ligase complex
WBGene00018620	<i>F48G7.9</i>	Positive regulation of glutamate-cysteine ligase activity
WBGene00003670	<i>nhr-80</i>	Nuclear Hormone Receptor family
WBGene00003877	<i>pept-1</i>	Peptide transporter family 1
WBGene00004259	<i>pyr-1</i>	CAD protein Glutamine-dependent carbamoyl-phosphate synthase Aspartate carbamoyltransferase Dihydroorotase
<i>TELCIR_14993</i>	NA	Glutamate-cysteine ligase catalytic subunit binding

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART. Those for which no *C. elegans* homologue could be found were identified by *H. contortus* GO term annotations, or if they had only an equivalent gene identified in PRJNA72569, they were identified by the PRJNA72569 GO term annotations.

**Table 6.6: Farm 2 predicted homologues annotated with GO terms associated with female related terms.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00000430	<i>ceh-5</i>	Homeobox protein <i>ceh-5</i>
WBGene00000837	<i>cul-2</i>	Cullin-2
WBGene00000868	<i>cyb-3</i>	G2/mitotic-specific cyclin-B3
WBGene00001377	<i>eya-1</i>	Eyes absent homolog 1
WBGene00002295	<i>let-19</i>	Mediator of RNA polymerase II transcription subunit 13
WBGene00003017	<i>lin-31</i>	Abnormal cell lineage protein 31
WBGene00003029	<i>lin-44</i>	Abnormal cell lineage protein 44
WBGene00003367	<i>mix-1</i>	Mitotic chromosome and X-chromosome-associated protein mix-1
WBGene00003638	<i>nhr-48</i>	Nuclear hormone receptor family member <i>nhr-48</i>
WBGene00004013	<i>pha-4</i>	Defective pharyngeal development protein 4
WBGene00004930	<i>sod-1</i>	Superoxide dismutase [Cu-Zn]
WBGene00006771	<i>tln-1</i>	Talin
WBGene00006768	<i>unc-32</i>	Probable V-type proton ATPase 116 kDa subunit a

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART.

**Table 6.7: Farm 3 predicted homologues annotated with GO terms associated with female related terms.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00003670	<i>nhr-80</i>	Nuclear Hormone Receptor family
WBGene00003794	<i>npp-8</i>	Nuclear Pore complex Protein
WBGene00007413	<i>pro-2</i>	Nucleolar complex protein 2 homolog

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART.

**Table 6.8: Farm 2 predicted homologues annotated with GO terms associated with male related terms.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00000837	<i>cul-2</i>	Cullin-2
WBGene00000868	<i>cyb-3</i>	G2/mitotic-specific cyclin-B3
WBGene00003029	<i>lin-44</i>	Abnormal cell lineage protein 44
WBGene00003058	<i>lov-1</i>	Location of vulva defective 1
WBGene00003605	<i>nhr-6</i>	Nuclear hormone receptor family member <i>nhr-6</i>
WBGene00007733	<i>smz-1</i>	Sperm Meiosis PDZ domain containing proteins
WBGene00020661	<i>smz-2</i>	Sperm Meiosis PDZ domain containing proteins

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART.

**Table 6.9: Farm 3 predicted homologues annotated with GO terms associated with male related terms.**

Gene stable ID	Gene name	Gene description or GO term
WBGene00004909	<i>snf-10</i>	Transporter

All genes were identified using GO terms returned by the BIOMART tool in WormBase ParaSite. *Caenorhabditis elegans* genes were the preferred choice for identifying function, and all *Teladorsagia circumcincta* PRJNA72569 equivalent genes and *Haemonchus contortus* homologues were first used to identify *C. elegans* homologues in BIOMART.

### 6.3.3 SNPs predicted to have high or moderate impact within regions of genetic differentiation

I hypothesised that genes containing SNPs that affected transcription of a protein or caused a codon change were more likely to be under IVM selection. In addition, I considered whether genes within a certain distance from a SNP might also be the SNP might also be under IVM selection. SnpEff (Cingolani et al., 2012), which estimates the effect of variants, was provided with variant sites of interest (as defined in 2.9.7) and the preliminary gene annotations. Only preliminary annotations are present on the Tci2 genome, and the length of untranslated regions (UTRs) for *T. circumcincta* genes are unknown. In addition, it is not known whether regions distant from a gene can regulate its transcription. For this reason, SNPs within 5 kb of a gene (SnpEff default setting), were considered to potentially modify a gene.

On Farm 2, 636,541 SNPs had  $F_{ST}$  values greater than the threshold of 0.116, and 1.13 million SNPs had a significant allele frequency change between the pre- and post-treatment populations ( $p < 0.05$ ). In total, 26,002 SNPs were retained which had an  $F_{ST}$  value greater than the threshold and a significant allele frequency difference. Of these, 676 were within the top 10 kb windows for Farm 2. Lastly, these 676 SNPs were filtered to obtain a list of SNP positions with mean depth of between 20-60x coverage in the Farm 2 data. This dataset of 212 SNPs was provided as input to SnpEff to obtain a list of SNPs, filtered by the above criteria, and which were also putatively predicted to have either a high or moderate impact phenotypically.

No SNPs were predicted to have a high impact on Farm 2. In total, four SNPs were predicted to have a moderate impact upon four *Tci2* genes, with modification of a fifth (Table 6.10). The equivalent PRJNA72569 genes and homologues in *H. contortus* and *C. elegans* are shown in Table 6.11 and Table 6.12. Each SNP was non-synonymous, introducing an amino acid change in the *Tci2* annotated proteins G540.T1, G10328.T1, G15772.T1 and G43444.T1, G43444.T2 (Table 6.10). *g540.t1* had no identifiable proteins in any of the three protein databases (Table 6.11). *g10328.t1* was predicted to be an aminopeptidase. The transcript *g15772.t1* was identified as *TELCIR\_03383*, a transposable element (TEs: reverse transcriptases, transposases, endo/exonucleases), with a homologue also identified in *H. contortus*. Lastly, *g43444* was identified as the DNA repair protein *rev-1* (Table 6.11). In addition, the SNP within this gene was predicted to be a downstream gene variant of the transcript *g43445.t1*. Note that this is arbitrary - any SNP within 5 kb of a gene will be annotated as either an upstream or downstream variant (Cingolani et al., 2012). This transcript had only one homologue identified, *HCON\_00177350*, a TE (Table 6.12).

**Table 6.10: Farm 2 moderate impact SNPs predicted by SnpEff, which were present within genetically differentiated 10 kb windows.**

Tci2 Contig	SNP	QUAL <sup>1</sup>	Annotation <sup>2</sup>	Tci2 Transcript
5	180615	999	Missense variant (g540.t1 protein_coding 1/4 c.23C>A p.Pro8His 23/348 23/348 8/115   )	<i>g540.t1</i>
140	229875	211	Missense variant (g15772.t1 protein_coding 1/1 c.328G>A p.Gly110Arg 328/570 328/570 110/189   )	<i>g15772.t1</i>
1305	28503	999	Missense variant (g43444.t1 protein_coding 3/12 c.2000G>A p.Arg667Gln 2000/2997 2000/2997 667/998   ) Missense variant (g43444.t2 protein_coding 3/12 c.2000G>A p.Arg667Gln 2000/2985 2000/2985 667/994   ) Downstream gene variant (g43445.t1 protein_coding   c.*3756G>A       3756   )	<i>g43444.t1</i> <i>g43444.t2</i> <i>g43445.t1</i>
5194	11514	999	Missense variant (g10328.t1 protein_coding 8/8 c.908T>C p.Ile303Thr 908/990 908/990 303/329   )	<i>g10328.t1</i>

Note that all SNPs were identified as having a high  $F_{ST}$  value, a significant allele frequency change between the pre- and post-ivermectin samples ( $p < 0.05$ ) and were sequenced to a mean depth of between 20-60x in the Farm 2 data. In addition they were present within 10 kb regions with  $F_{ST}$  values greater than the threshold. <sup>1</sup>QUAL reflects the confidence in the variance seen at the SNP position ( $-\log_{10}$ [probability that the alternate allele is wrongly called]).

<sup>2</sup>Annotations include: variant type, transcript, transcript-type, exon-with-SNP/total-exons, HGVS.c - the SNP and it's position within the cDNA sequence, HGVS.p - the amino acid change and it's position within the protein sequence, cDNA position/total cDNA length, protein position/total protein sequence length, distance to first/last exon if upstream/downstream variant.

Table 6.11: Farm 2 gene transcripts containing SNPs of moderate impact.

Tci2 Transcript ID	<i>T. circumcineta</i> <sup>1</sup> PRJNA72569 gene	<i>T. circumcineta</i> <sup>1</sup> PRJNA72569 genomic location <sup>2</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic location <sup>4</sup>	<i>C. elegans</i> <sup>5</sup> gene	<i>C. elegans</i> <sup>5</sup> Gene stable ID	<i>C. elegans</i> <sup>5</sup> genomic location <sup>4</sup>	Gene description
g540.t1	NA	NA	NA	NA	NA	NA	NA	Unknown
g10328.t1	TELCIR_08194	1466:52385-62418	HCON_00142190	V:15962912-16009730	Y42A5A.1	WBGene00012776	V:11088154-11097970	Amino-peptidase
g15772.t1	TELCIR_03383	265:87651-88220	HCON_00170720	X:11864855-11865637	NA	NA	NA	TE
g43444.t1	TELCIR_02997	210:159381-163463	HCON_00104960	IV:11508359-11509732	ZK675.2 (rev-1)	WBGene00014066	II:7900647-7904636	DNA repair protein REV1
g43444.t2	TELCIR_02997	210:159381-163463	HCON_00104960	IV:11508359-11509732	ZK675.2 (rev-1)	WBGene00014066	II:7900647-7904636	DNA repair protein REV1

<sup>1</sup>*Teladorsagia circumcineta* <sup>2</sup>Contig:nucleotide range <sup>3</sup>*Haemonchus contortus* <sup>4</sup>Chromosome:nucleotide range <sup>5</sup>*Caenorhabditis elegans*. TE = Transposable element.



Table 6.12: Farm 2 gene transcripts putatively impacted by a SNP of moderate impact in a nearby gene.

Tci2 Transcript ID	<i>T. circumcineta</i> <sup>1</sup> PRJNA72569 gene	<i>T. circumcineta</i> <sup>1</sup> PRJNA72569 genomic location <sup>2</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic location <sup>4</sup>	<i>C. elegans</i> <sup>5</sup> gene	<i>C. elegans</i> <sup>5</sup> Gene stable ID	<i>C. elegans</i> <sup>5</sup> genomic location <sup>4</sup>	Gene description
g43445.t1	NA	NA	<i>HCON_00177350</i>	X:23863014-23865806	NA	NA	NA	TE

<sup>1</sup>*Teladorsagia circumcineta* <sup>2</sup>Contig:nucleotide range <sup>3</sup>*Haemonchus contortus* <sup>4</sup>Chromosome:nucleotide range <sup>5</sup>*Caenorhabditis elegans*. TE = Transposable element.

In total, 24,052 SNPs of high  $F_{ST}$  value were retained following filtering of the Farm 3 data (see 2.9.4). Twice this number of SNPs (52,426) had a significant difference in allele frequency between the pre- and post-IVM samples ( $p < 0.05$ ). Eleven SNPs were present in both datasets. Of these, two were found to reside within the genetically differentiated windows identified on Farm 3. Both had coverage depth between 20 and 60x but neither were deemed to have high or moderate impact by SnpEff.

### 6.3.4 Comparison of results from both farms identified some common regions potentially under selection by IVM

Sixteen genetically differentiated windows were identified on both farms, containing 19 Tci2 genes (Table 6.13). Two of these 16 windows had no annotated genes, although one of these, on Tci2 contig 2028, was adjacent to a differentiated window that did contain a gene. Six of these genes were identified as TEs (Table 6.14). Additional predicted proteins, identified by BLASTP search of the Tci2 amino acid sequences, were common to both farm analyses - yet the Tci2 genomic regions were different in each farm analysis. This could indicate that the two (or more) Tci2 genes could represent different genes within the same gene family, or paralogues. By identifying genes under selection on both farms in this way, it may be possible to determine gene families potentially important in IVM sensitivity on both farms. Alternatively, a gene under IVM selection on both farms may be represented by two or more allelic sequences in the draft Tci2 genome. Read alignment to these alleles could be different on each farm. For this reason, protein lists generated by BLASTP searches of *C. elegans*, *H. contortus* and the *T. circumcincta* PRJNA72569 protein databases were compared between farms. A further ten proteins potentially under selection by IVM on both farms were identified in this way (Table 6.15). These included PGP-3, which was identified as the homologue of four separate Tci2 proteins within genetically differentiated windows. Eleven Tci2 genes, identified in the Farm 2 analysis, were predicted to be the same PRJNA72569 equivalent gene, a TE. Three of the proteins identified on both farms were predicted to be peptide transporters (PEPT-1, PEPT-2 and PEPT-3). Both *Cel-pept-2* and *Hco-pept-3* were homologues of the same Tci2 gene.

**Table 6.13: Numbers of Tci2 annotated genes and predicted proteins within genetically differentiated regions which were identified in both farm analyses.**

	Number detected on both farms
10 kb windows above the $F_{ST}$ threshold (Tci2 contigs)	16 (14)
Tci2 genes identified within these windows (transcripts)	19 (21)
<i>Teladorsagia circumcincta</i> PRJNA72569 reference genome proteins predicted (% of all Tci2 genes)	9 (47.4%)
<i>Haemonchus contortus</i> proteins predicted (% of all Tci2 genes)	15 (73.7%)
<i>Caenorhabditis elegans</i> proteins predicted (% of all Tci2 genes)	6 (31.6%)
Tci2 genes for which proteins were predicted in all three datasets (% of all Tci2 genes)	4 (21.1%)
Tci2 genes for which parasite proteins were predicted, without a predicted homologue in <i>Caenorhabditis elegans</i> (% of all Tci2 genes); putatively parasite-specific genes	9 (47.4%)
Tci2 genes for which proteins were identified solely in the PRJNA72569 database (% of all Tci2 genes); putatively <i>Teladorsagia circumcincta</i> -specific genes	0 (0.0%)
Tci2 genes for which no protein was identified (% of all Tci2 genes)	4 (21.1%)
10 kb windows with no Tci2 gene (% of all windows)	2 (12.5%)

**Table 6.14: 10 kb windows of the *Teladorsagia circumcincta* Tci2 draft genome identified on both Farm 2 and Farm 3 with  $F_{ST}$  values greater than all comparison thresholds.**

Tci2 genomic position <sup>1</sup>	Tci2 transcript name	<i>T. circumcincta</i> <sup>2</sup> PRJNA72569 gene	<i>T. circumcincta</i> <sup>2</sup> PRJNA72569 genomic position <sup>1</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic position <sup>4</sup>	<i>C. elegans</i> <sup>5</sup> gene	<i>C. elegans</i> <sup>5</sup> genomic position <sup>4</sup>	Gene description
4:1220000-1230000	<i>g522.t1</i> <i>g522.t2</i>	TELCIR_11371	2975:898-54044	HCON_00015810	I:22624698-22655013	F36H2.3	I:9251838-9261657	Complement C3d receptor 2
41:780000-790000	<i>g2978.t1</i>	TELCIR_13612	4637:257-937	HCON_00047360	II:19224350-19224700	NA	NA	Unknown (single exon)
41:780000-790000	<i>g2977.t1</i>	TELCIR_10574	2485:16509-36197	HCON_00168150	X:7239913-7241341	NA	NA	TE
41:780000-790000	<i>g2976.t1</i>	TELCIR_01154	47:11070-11630	HCON_00187972	X:40329402-40335075	NA	NA	TE
163:800000-810000	<i>g16889.t1</i>	NA	NA	HCON_00193670	II:41464399-41483472	NA	NA	Unknown
382:270000-280000	<i>g24918.t1</i>	TELCIR_03613	307:55728-119232	NA	NA	ZK20.6 ( <i>nep-1</i> )	II:11654214-11657172	Neprilysin-1
703:100000-110000	<i>g33104.t1</i>	NA	NA	HCON_00193210	II:30974948-30990518	T22C8.2 ( <i>chhy-1</i> )	II:8614546-8617603	Hyaluronidase
970:160000-170000	<i>g38597.t1</i>	NA	NA	HCON_00174350	X:18160059-18160850	NA	NA	TE
1041:30000-40000	<i>g39761.t1</i>	NA	NA	HCON_00052580	II:28633156-28635145	NA	NA	TE
1359:160000-170000	<i>g44198.t1</i>	NA	NA	NA	NA	NA	NA	NA

Tci2 genomic position <sup>1</sup>	Tci2 transcript name	<i>T. circumcincta</i> <sup>2</sup> PRJNA72569 gene	<i>T. circumcincta</i> <sup>2</sup> PRJNA72569 genomic position <sup>1</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic position <sup>4</sup>	<i>C. elegans</i> <sup>5</sup> gene	<i>C. elegans</i> <sup>5</sup> genomic position <sup>4</sup>	Gene description
1476:90000-100000	<i>g45559.t1</i>	TELCIR_07851	1337:8475-27067	HCON_00025940	I:36693463-36695728	K04G2.1 ( <i>eif-2B</i> )	I:8026606-8028752	Eukaryotic translation initiation factor 2 subunit 2
476:90000-100000	<i>g45560.t1</i>	NA	NA	HCON_00000560	I:665765-672922	NA	NA	TE
1476:90000-100000	<i>g45562.t1</i>	NA	NA	HCON_00066030	III:1736614-1737678	NA	NA	TE
1476:90000-100000	<i>g45561.t1</i>	NA	NA	NA	NA	NA	NA	NA
1646:20000-30000	<i>g47181.t1</i>	NA	NA	NA	NA	NA	NA	NA
2028:50000-60000	<i>g50265.t1</i>	TELCIR_14323 <sup>6</sup>	5289:2069-9396	HCON_00063720	II:44978432-44996563	NA	NA	NAB region 2 domain
2028:70000-80000	<i>g50266.t1</i>	TELCIR_01040	42:20306-32459	HCON_00067270	III:3717209-3720584	T28B4.3 ( <i>ttr-6</i> )	X:6596425-6597121	Transthyretin-Related family domain
2028:80000-90000	NA	NA	NA	NA	NA	NA	NA	NA
2170:30000-40000	<i>g51079.t1</i> <i>g51079.t2</i>	TELCIR_07563	1244:13623-84320	HCON_00010920	I:14733034-14747180	C06G8.2 ( <i>pept-2</i> )	IV:10790598-10794597	Peptide transporter family
6280:0-10000	NA	NA	NA	NA	NA	NA	NA	NA

Tci2 genomic position <sup>1</sup>	Tci2 transcript name	<i>T. circumcincta</i> <sup>2</sup> PRJNA72569 gene	<i>T. circumcincta</i> <sup>2</sup> PRJNA72569 genomic position <sup>1</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic position <sup>4</sup>	<i>C. elegans</i> <sup>5</sup> gene	<i>C. elegans</i> <sup>5</sup> genomic position <sup>4</sup>	Gene description
6442:0-10000	<i>g11819.t1</i>	NA	NA	NA	NA	NA	NA	NA

Included are those homologues and PRJNA72569 equivalent genes identified by BLASTP of the translated nucleotide sequence. These were identified using the WSI *farm\_blast*. If no homologue was identified by this BLASTP for the Tci2 translated nucleotide then NA is written. <sup>1</sup>Contig:nucleotide range. <sup>2</sup>*Teladorsagia circumcincta* <sup>3</sup>*Haemonchus contortus* <sup>4</sup>Chromosome:nucleotide range. <sup>5</sup>*Caenorhabditis elegans*. <sup>6</sup>Identified by Choi et al. 2017 in a *T. circumcincta* multi-drug resistance genome wide association study (see section 6.3.5.1). TE = Transposable element.

**Table 6.15: Proteins predicted to be homologues or equivalent PRJNA72569 proteins of Tci2 genes located within separate regions of differentiation in each farm analyses.**

<i>C. elegans</i> <sup>1</sup> gene	<i>C. elegans</i> <sup>1</sup> gene stable ID	<i>C. elegans</i> <sup>1</sup> genomic location <sup>2</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic location <sup>2</sup>	<i>T. circumcincta</i> <sup>4</sup> PRJNA72569 gene	<i>T. circumcincta</i> <sup>4</sup> PRJNA72569 genomic location <sup>5</sup>	Gene description
<i>C09H10.8 (glb-4)</i>	WBGene00007502	II:11109584-11111169	<i>HCON_00063260</i>	II:44063590-44076404	<i>TELCIR_01028</i>	40:407974-452373 <sup>6</sup>	Globin related
<i>D2085.1 (pyr-1)</i>	WBGene00004259	II:8651109-8658808	<i>HCON_00008370</i>	I:10891644-10913956	NA	NA	CAD protein Glutamine-dependent carbamoyl-phosphate synthase Aspartate carbamoyltransferase Dihydroorotase
<i>T20B3.7 (phy-3)</i>	WBGene00004026	V:16833316-16839141	NA	NA	NA	NA	Proline Hydroxylase
<i>E01B7.1 (usp-50)</i>	WBGene00008441	V:17924916-17934225	<i>HCON_00144310</i>	V:19589136-19603459	<i>TELCIR_07375</i>	1183:11919-19623	Ubiquitin Specific Protease
<i>ZK455.7 (pgp-3)</i>	WBGene00003997	X:11352136-11357094	<i>HCON_00042800</i>	II:11890313-11916024	NA	NA	Multidrug resistance protein <i>pgp-3</i>
NA	NA	NA	<i>HCON_00050050</i>	II:24598380-24658474	NA	NA	WD40 repeat domain ( <i>tcab-1</i> )
NA	NA	NA	<i>HCON_00121860</i>	IV:37733818-37736775	<i>TELCIR_07666</i>	1274:16307-16852	TE

<i>C. elegans</i> <sup>1</sup> gene	<i>C. elegans</i> <sup>1</sup> gene stable ID	<i>C. elegans</i> <sup>1</sup> genomic location <sup>2</sup>	<i>H. contortus</i> <sup>3</sup> gene	<i>H. contortus</i> <sup>3</sup> genomic location <sup>2</sup>	<i>T. circumcincta</i> <sup>4</sup> PRJNA72569 gene	<i>T. circumcincta</i> <sup>4</sup> PRJNA72569 genomic location <sup>5</sup>	Gene description
NA	NA	NA	<i>HCON_00133270</i>	V:2943852-2948958	NA	NA	Oxidoreductase (C14E2.4)
NA	NA	NA	<i>HCON_00176880</i>	X:22986334-22988317	<i>TELCIR_12769</i>	3936:2914-3336	TE
NA	NA	NA	<i>HCON_00178645</i>	X:26119795-26179301	NA	NA	Peptide transporter family ( <i>pept-1</i> )

Homologues and equivalent proteins were predicted for Tci2 genes within genetically differentiated regions on each farm by BLASTP of the translated nucleotide. Shown in this table are homologues/equivalent genes common to both farm analyses, but which were identified using the WSI *farm\_blast* for a different Tci2 gene in each farm analysis. If no homologue was identified by this BLASTP for the Tci2 translated nucleotide then NA is written.

<sup>1</sup>*Caenorhabditis elegans* <sup>2</sup>Chromosome:nucleotide range <sup>3</sup>*Haemonchus contortus* <sup>4</sup>*Teladorsagia circumcincta* <sup>5</sup>Contig:nucleotide range. <sup>5</sup>*C. elegans*. <sup>6</sup>Choi et al. (2017) identified four SNPs approximately 200,000 bp distant from this gene (using ddRAD-Seq) which were potentially under selection by IVM (see section 6.3.5.1). TE = Transposable element.



### 6.3.5 Comparison of results with other GWAS

Four GWAS investigating IVM resistance in parasitic nematodes have been carried out to date. Genes identified in the present study were compared to data from prior studies with the purpose of identifying common genes found to be under selection by IVM across studies, isolates and species. A *T. circumcincta* study using Pool-Seq to compare a susceptible strain and a multidrug resistant strain (Choi et al., 2017), was considered first. Two *H. contortus* studies were also considered, one using WGS to compare individuals of geographically separated field isolates (Sallé et al., 2019), and the other using Pool-Seq and RNA-Seq to investigate IVM resistance in a genetic cross (BUG Consortium, unpublished). Lastly, a study investigating IVM resistance using clinical isolates of the filarial nematode, *O. volvulus*, was considered (Doyle et al., 2017). For each GWAS, genes compared were those identified by the authors as being of potential interest.

#### 6.3.5.1 *T. circumcincta* New Zealand study

A NZ study produced an inbred susceptible strain ( $S_{\text{inbred}}$ ), using an anthelmintic susceptible lab isolate, originally obtained in the 1950s before widespread anthelmintic use (Choi et al., 2017). In 1996, a triple resistant (BZ, LEV and IVM) population was isolated from the field and this was passaged for five generations, treating each generation with all three anthelmintics simultaneously to generate a strain that was highly resistant to all three drugs. Subsequently, this strain was crossed with the  $S_{\text{inbred}}$  strain via surgical transfer of approximately 300 immature female susceptible worms and 300 mature adult male resistant worms. In total, three surgical crosses interspersed by interbreeding of siblings for five generations each, and multidrug selection (BZ, LEV and IVM) were performed to produce the triple resistant  $RS^3$  strain. The authors then performed two experiments. Firstly, they compared the  $S_{\text{inbred}}$  to the multi-drug treated  $RS^3$  strain using Pool-Seq of mixed-sex adults (300-500 worms per pool). They also performed ddRAD-Seq using the F2 generation of the first intercross ( $F_2^1$ ). For this, two groups, each of nine goat kids, were infected with  $F_2^1$  and one group was treated with IVM. From these 18 goats, 24 IVM

treated male *T. circumcincta* and 24 untreated males were selected for ddRAD-Seq.

The Choi study identified 18 genes of interest that were present in both their Pool-Seq and ddRAD-Seq experiments. However, none of these were identified in the present study. The comparison was extended to include all genes within highly differentiated regions (z-score 4.5) in the introgression Pool-Seq analysis, which also contained a ddRAD-Seq locus (of any score) on the same contig (Choi et al. (2017) table S13). Just one gene (*TELCIR\_01265*) within Choi's study was also predicted to be under selection by IVM on Farm 2, with no genes identified in the Farm 3 data. This gene does not appear to have a *C. elegans* homologue; it has a macin domain and may be involved in the defense response.

A comparison of SNP data was also undertaken. ddRAD-Seq SNPs with data from at least ten individuals in each sample population, and with an  $F_{ST}$  z-score of 2 or greater (the lowest score reported by Choi; SNPs with no z-score reported were discarded) were selected, providing 149 SNPs to compare with the present study. As Choi et al. had aligned their data to the PRJNA72569 genome, the SNPs could not be directly compared to the Tci2 data. Instead, these 149 SNPs were compared with the PRJNA72569 contigs containing the equivalent Tci2 genes identified in each farm analysis. Both farms had just one PRJNA72569 gene containing a ddRAD-Seq SNP. This gene was *TELCIR\_14323* on PRJNA72569 contig 5289, the equivalent gene of *g50265* on Tci2 contig 2028 (Table 6.14). This gene is predicted to have transcriptional regulatory activity and contains a NAB co-repressor domain. The *C. elegans* homologue of this gene, *mab-10*, is widely expressed in the worm, and is involved in the moulting cycle (Harris and Horvitz, 2011) and reproductive morphology (Hodgkin, 1983). Although no other ddRAD-Seq SNPs were identified within PRJNA72569 genes on either farm, ddRAD-Seq SNPs on PRJNA72569 contigs shared between the studies were identified (Appendix 12). Of these, three PRJNA72569 contigs were of particular interest. In total, four ddRAD-Seq SNPs were identified between PRJNA72569 contig 40:220119-220176, roughly 200,000 bp from *TELCIR\_01028*, predicted to be the homologue of the *C. elegans* gene *glb-4*, identified in both farm analyses (Table 6.15). Of the four SNPs, all had a z-score >2.5, with one displaying a high z-score >4.5. Secondly, a ddRAD-Seq SNP on PRJNA72569 contig 1274 was identified.

This SNP was 3859 bp from the gene *TELCIR\_07666*, which is a single exon of unknown function (most likely to be a TE), and was also identified in both farm datasets. Lastly, two ddRAD-Seq SNPs were present on PRJNA72569 contig 43. In the Farm 2 dataset, four PRJNA72569 equivalent genes were identified on this contig with the following *C. elegans* homologues; *mpz-1*, *mnk-1*, *rhgf-2* and *lgc-27*. The closest gene to a SNP was *Tci-rhgf-2*, approximately 49 kb away.

### 6.3.5.2 *H. contortus* studies

#### Comparisons of *H. contortus* samples from across the globe

A very interesting analysis comparing *H. contortus* isolates across and within continents was reported by Sallé et al. (2019). As part of this work, WGS of IVM susceptible (NAM, ZAI and AUS.2) and IVM resistant (STA.1, STA.3 and AUS.1) isolates were compared within continent by pairwise XP-CLR analysis. This is a cross-population composite likelihood ratio test. Both the allele frequency differentiation relative to the reference population (under assumptions of neutrality within the reference population), and the extent of LD are combined to locate regions of recent selection within the genome (Chen et al., 2010). Genes of interest within 2000 bp of a selection hit were searched for, with six genes identified as putatively under selection by IVM: *aex-3*, *B0361.4*, *glc-4*, *pgp-11*, *snf-9* and *unc-24* (Sallé et al., 2019). None of these genes appeared in either the Farm 2 or Farm 3 analyses, but members of the same gene classes were detected. *pgp-3* appeared to be under selection on both farms, while *snf-3* and *snf-10*, members of the neurotransmitter:sodium symporter family, were identified on Farms 2 and 3 respectively. On Farm 2 only, the GluCl gene, *glc-3*, and the gene *unc-32*, were identified as potentially under selection by IVM.

#### *H. contortus* genetic cross (BUG Consortium)

A study using a laboratory cross between the drug-susceptible *H. contortus* isolate MHco3(ISE) and the triple resistant isolate MHco18(UGA), resistant to BZ, LEV and IVM, identified a large region of differentiation by  $F_{ST}$  on Chromosome V associated with IVM selection (Laing, *pers comm*). This region was in two main peaks: V:37-38 Mbp and V:45-48 Mbp. Corresponding regions had been detected by Doyle et al. (2019) in two laboratory backcrosses under IVM selection

(MHco3(ISE)xMHco4(WRS) and MHco3(ISE)xMHco10(CAVR)) and also by Sallé et al. (2019) as described above. The BUG Consortium generated a genetic cross between female MHco3(ISE) and male MHco18(UGA); F1 progeny were crossed with their siblings to provide the F2 generation, which was treated with IVM. Pool-Seq was performed on 200 F3 L3 obtained pre- and post-IVM treatment of the F2 generation. In addition to this, RNA-Seq was performed using the F2 IVM-treated adult worms and untreated F2 controls. *H. contortus* genes were compared with *H. contortus* homologues of Tci2 genes in the present study.

On Farm 2, five Tci2 genes were identified as homologues of four *H. contortus* genes located within the Chromosome V region identified by the BUG consortium (Table 6.16). The *H. contortus* gene *HCON\_00154992* is a predicted 7TM-GPCR serpentine receptor of class R with no *C. elegans* homologue. *HCON\_00155250* is a homologue of two Tci2 genes (on separate Tci2 contigs) within the Farm 2 data. It is predicted to encode an uncharacterised oxidoreductase, with a *C. elegans* homologue *E02C12.19*. *HCON\_00155250* was downregulated in expression in both male (-0.73 log<sub>2</sub> fold change, p<0.001) and female (-1.09 log<sub>2</sub> fold change, p<0.01) IVM treated *H. contortus* (Laing, *pers comm*). A third gene, *HCON\_00161360* is a homologue of the *C. elegans* gene, *phat-1*. *phat-1* contains a ShKT domain and is expressed in the pharyngeal gland cells, which are thought to be equivalent to mammalian salivary glands (Smit et al., 2008). The last gene identified in both studies was *HCON\_00162680*, the homologue of *ape-1*, a gene involved in negative regulation of p53-induced apoptosis in *C. elegans* (Bergamaschi et al., 2003). No genes identified within the Farm 3 data were also found within the Chromosome V region identified by the BUG Consortium.

**Table 6.16: Genes shared between Farm 2 and Chromosome V peak of the BUG Consortium *Haemonchus contortus* genetic cross.**

Tci2 genomic location <sup>1</sup>	Tci2 transcript	<i>Haemonchus contortus</i> gene	<i>Haemonchus contortus</i> genomic location <sup>2</sup>	<i>Caenorhabditis elegans</i> gene stable ID	<i>Caenorhabditis elegans</i> gene	Gene description	Differentially expressed? <sup>3</sup>
441:0-10000 <sup>4</sup>	<i>g26855.t1</i>	<i>HCON_00154992</i>	V:37059054-37061784	NA	NA	7TM-GPCR, serpentine receptor class R (Str)	No
30:620000-640000	<i>g2278.t1</i>	<i>HCON_00155250</i>	V:37333743-37339458	WBGene00271797	<i>E02C12.19</i>	Uncharacterised oxidoreductase Dhs-27	Yes; males and females (p<0.01)
708:120000-140000	<i>g33217.t1</i>	<i>HCON_00155250</i>	V:37333743-37339458	WBGene00271797	<i>E02C12.19</i>	Uncharacterised oxidoreductase Dhs-27	Yes; males and females (p<0.01)
357:310000-320000	<i>g24147.t1</i>	<i>HCON_00161360</i>	V:45284232-45287187	WBGene00016732	<i>C46H11.8 (phat-1)</i>	Pharyngeal gland Toxin-related, ShKT domain protein	No
83:80000-90000	<i>g12991.t1</i>	<i>HCON_00162680</i>	V:47133046-47147156	WBGene00000146	<i>F46F3.4 (ape-1)</i>	Apoptotic enhancer 1 protein	No

*H. contortus* homologues located within regions of genetic differentiation in the Farm 2 data were compared with *H. contortus* genes identified within a region of high  $F_{ST}$  on Chromosome V following Pool-Seq comparing progeny of F2 treated adults pre- and post-IVM. <sup>1</sup>Contig:nucleotide range. <sup>2</sup>Chromosome:nucleotide range. <sup>3</sup>RNA-Seq for differential expression of genes between F2 generation *H. contortus* adult worms treated with IVM and untreated controls. Males and females were compared separately. <sup>4</sup>This window is rank 1/576 in Farm 2 data. However, Tci2 contig 441 is not present in the Farm 3 data.

Following comparison with the Chromosome V region, data were compared with the *H. contortus* BUG Consortium RNA-Seq experiment. Genes found to be significantly differentially expressed within the RNA-Seq dataset (3660 genes,  $p < 0.05$ ) were compared with the genes within high  $F_{ST}$  windows in this study. A total of 106 *H. contortus* genes, with differential expression values ranging from -3.21 to 4.44  $\log_2$  fold change, were identified as homologues of Tci2 genes within the Farm 2 dataset (Appendix 13). Noteworthy was reduced expression of *Hco-pgp-3* in IVM-selected *H. contortus* males (-0.72  $\log_2$  fold change) and females (-1.45  $\log_2$  fold change) compared with the IVM naïve controls. A K<sup>+</sup>/Cl<sup>-</sup> Co-transporter, *kcc-1*, showed increased expression in IVM selected *H. contortus* males (0.22  $\log_2$  fold change). Several genes involved in metabolism pathways were also present, in both datasets, including some involved in lipid metabolism.

Of the genes shared by Farm 2, the third greatest increase in expression (by decreasing rank order), in IVM selected *H. contortus* adults was a LGIC, *lgc-42* (2.14  $\log_2$  fold change in males). An acetylcholine receptor (AChR) subunit, *acr-17*, also showed increased expression in IVM selected *H. contortus* males (1.03  $\log_2$  fold change). Only two *H. contortus* genes showed increased expression greater than *lgc-42* in IVM selected adults relative to controls; *HCON\_00142620*, a nematode gene of unknown function, putatively a peptidase, and *HCON\_00102660*, also of unknown function (4.08 and 4.44  $\log_2$  fold change respectively, males). The gene with the greatest down-regulation in IVM selected adults in the Farm 2 data was the predicted homologue of *ZC155.4*, a gene involved in lipid metabolism (-3.21  $\log_2$  fold change, females). Another gene, a complement receptor, previously identified on both Farm 2 and Farm 3 in the  $F_{ST}$  analysis, *F36H2.3*, was also reduced in IVM selected females (-1.98  $\log_2$  fold change) but was increased in males (0.37  $\log_2$  fold change).

Farm 3 had 12 *H. contortus* homologues identified as differentially expressed between the *H. contortus* F2 generation IVM treated and untreated adults. These are shown in Appendix 13. The gene with the greatest reduction in expression in the IVM selected females was *F36H2.3*, followed by *pgp-3*. Those with increased expression included *snf-10*, a member of the sodium:neurotransmitter symporter family (0.44  $\log_2$  fold change, males), and *lev-9*, important in clustering of AChRs at synapses (Gendrel et al., 2009), (0.70  $\log_2$  fold change, males). The gene with

the greatest increase in expression was a protein tyrosine phosphatase, *T28F4.3*, with 0.85 log<sub>2</sub> fold change in *H. contortus* males.

### 6.3.5.3 *Onchocerca volvulus*

Doyle et al. (2017) investigated the genome of the human filarial nematode *O. volvulus* for regions under selection between sub-optimal responders (SOR) and good responders (GR) to IVM treatment from clinical subjects in Ghana and Cameroon. From their QTL analysis, 403 *O. volvulus* genes were identified as being within regions of genetic differentiation between the SOR and GR worms. Using the BIOMART tool in WormBase ParaSite, homologues of these genes were identified in *C. elegans*, *H. contortus* and *T. circumcincta* PRJNA72569 genomes. In total, homologues were identified for 229 genes (Appendix 14).

Farm 2 had five Tci2 genes shared with the dataset of Doyle et al. (2017) (Table 6.17). These included two genes with *H. contortus* homologues (*HCON\_00142180* and *HCON\_00142190*), which lie adjacent to each other on Chromosome V, between 15.93-16.01 Mbp. Farm 3 had only a single gene with a homologue identified within the *O. volvulus* dataset (Table 6.18). This was a TE, which in *H. contortus* was positioned within an intron of the gene *HCON\_00191600*, a dehydratase, a homologue of *C. elegans* *R107.2*.

Doyle et al. (2017) also categorised genes as being within pathways of interest potentially related to IVM sensitivity. Each farm dataset was manually searched for these genes, using the descriptive part of the *C. elegans* gene name (e.g. 'unc') to find members of the same gene class. These included those associated with (i) IVM, (ii) neurotransmission, (iii) LIN-12/Notch signalling, (iv) muscle assembly/myosin organisation and (v) lipid synthesis and storage/stress. Nine Farm 2 genes were identified, one of which was identified previously in the *O. volvulus* comparisons (*nrf1-1*, Table 6.17). The other eight were in the same gene class as *O. volvulus* genes and included *k1p-18*, *unc-32*, *acs-7*, *tat-4* and four LGIC (Table 6.19). However, the function of these genes may not necessarily correspond to the same descriptions as provided for the *O. volvulus* genes by Doyle et al. There were no equivalent genes identified in the Farm 3 analysis.

**Table 6.17: Genes within genetically differentiated regions in Farm 2 data with potential homologues also identified in an *Onchocerca volvulus* ivermectin study.**

Tci2 transcript	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> stable ID	Gene Description
<i>g10328.t1</i>	TELCIR_08194*	HCON_00142190*	Y42A5A.1	Metalloaminopeptidase activity
<i>g26501.t1</i> <i>g26501.t2</i>	TELCIR_13613	HCON_00119540	C01F6.6 ( <i>nrf1-1</i> ) <sup>1</sup>	Mammalian Na/H exchange regulatory factor like
<i>g5596.t1</i>	TELCIR_00201	HCON_00135630	F25B3.3 ( <i>rgef-1</i> )	Rap Guanine nucleotide Exchange Factor homolog
<i>g37707.t1</i>	TELCIR_12798	HCON_00142180	Y42A5A.4 ( <i>cdkl-1</i> )	Cyclin-dependent kinase-like 1
<i>g42704.t1</i>	TELCIR_10244	HCON_00111520	Y56A3A.13 ( <i>nft-1</i> )	Nitrilase and fragile histidine triad fusion protein NitFhit Bis(5'-adenosyl)-triphosphatase Nitrilase homolog

Homologues of genes identified within genomic regions of high differentiation between sub-optimal responders and good responders of female *O. volvulus* filarial worms treated with ivermectin were compared with genes identified on Farm 2. An asterisk (\*) indicates that a different homologue was identified by BIOMART. The gene identified by a BLASTP search of the Tci2 translated nucleotide sequence is provided here. <sup>1</sup>Note that this gene was within a pathway of interest identified by Doyle et al. as containing more than one gene within QTL regions (see Table 6.19)

**Table 6.18: Genes within genetically differentiated regions in Farm 3 data with potential homologues also identified in an *Onchocerca volvulus* ivermectin study.**

Tci2 transcript	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> stable ID	Gene Description
<i>g34923.t1</i>	TELCIR_25085	HCON_00192370	NA	TE

Homologues of genes identified within genomic regions of high differentiation between sub-optimal responders and good responders of female *O. volvulus* filarial worms treated with ivermectin were compared with genes identified on Farm 3. TE = Transposable element.



**Table 6.19: Farm 2 *Caenorhabditis elegans* homologues with similar description to *Onchocerca volvulus* genes.**

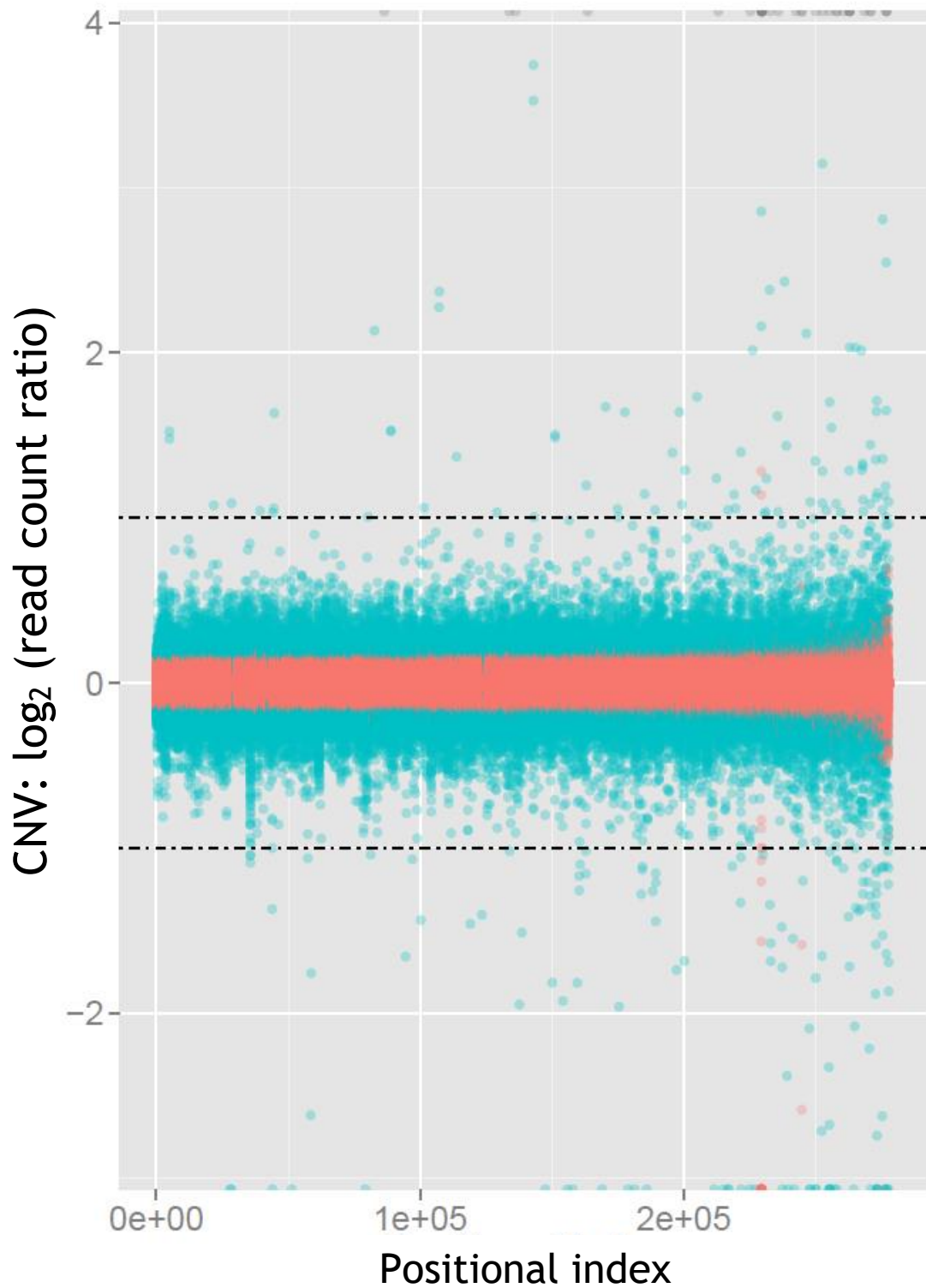
<i>Caenorhabditis elegans</i> gene (this study)	<i>Onchocerca volvulus</i> gene (Doyle et al., 2017)	Pathway/Functional characteristic
<i>klp-18</i>	<i>klp-11</i>	IVM-associated
<i>lgc-27</i> <i>lgc-39</i> <i>lgc-41</i> <i>lgc-42</i>	<i>lgc-46</i> <i>lgc-47</i>	Neurotransmission
<i>unc-32</i>	<i>unc-44</i>  <i>unc-17</i> <i>unc-31</i> <i>unc-26</i>  <i>unc-82</i>	IVM-associated  Neurotransmission  Muscle assembly/myosin organisation
<i>nrfl-1</i>	<i>nrfl-1</i>	Neurotransmission
<i>acs-7</i> <i>tat-4</i>	<i>acs-16</i> <i>tat-2</i>	Lipid synthesis and storage/stress

The pathway or functional characteristic is given for each of the *O. volvulus* genes. These genes were identified by Doyle et al. within QTL peaks and were predicted to have functional characteristics or pathways in common with other genes identified in QTLs in the data from Ghana or Cameroon. Genes of the same ‘gene class’ were identified in the Farm 2 data. Nevertheless, it is important to note that these genes may not have the same function as those identified by Doyle et al. 2017.

### 6.3.6 P-glycoproteins did not show copy number variation between pre- and post-treatment samples

Differences in copy number of genes or genomic regions can arise between individuals or populations following duplication, or deletion, of a gene or genomic region within a genome. If whole genes are duplicated, these are known as paralogues, and it has been suggested that increased copy number of *Tci-pgp-9* is associated with IVM resistance (Choi et al., 2017; Turnbull et al., 2018). In contrast, deletion of genes or genomic regions has been associated with BZ (Kwa et al., 1993) and LEV (Barrère et al., 2014) resistance. To investigate whether any copy number variants (CNV) were associated with IVM resistance on these farms, the program CNV-Seq (Xie and Tammi, 2009), was used to compare aligned read counts between the pre- and post-IVM treatment samples for each farm (Figure 6.1). As CNV is less accurately determined using Pool-Seq data than for individual sequencing, data were filtered as described in 2.9.9. For each comparative analysis, the post-IVM sample was compared to the pre-IVM sample. Hence, if the  $\log_2$  (read count ratio) is positive, read counts are higher in the

post-IVM sample compared to the pre-IVM sample from the same farm. It is important to note that the mean read counts for all post-IVM samples were significantly lower ( $p < 0.05$ ) than for all pre-treatment samples (Table 6.20).



**Figure 6.1: Farm 2 copy number variation.** Each point represents one 5 kb window. Those coloured in blue are significant at  $p < 0.00001$ . The dashed lines indicate  $\log_2$  (read count ratio) of 1 and -1. Farm 3 pairwise comparisons were similar in appearance.

**Table 6.20: Mean and median read counts per 5000 bp window for each sample.**

Sample	Mean read count per window	Median read count per window
Farm 2 Pre	2983	2689
Farm 2 Post	2880	2592
Farm 3 Pre_A	3100	2803
Farm 3 Pre_B	3033	2741
Farm 3 Post_A	2759	2494
Farm 3 Post_B	2793	2523

**Read counts reported by CNV-Seq after first filtering out unmapped reads.**

No deletions were identified in the post-IVM samples following filtering. A narrow, 98 bp intronic region within a cation transmembrane transporter gene was identified to have increased CNV on Farm 2 ( $\log_2(\text{read count ratio})$  3.6). However, apart from this stack of reads, the region generally had few reads mapping, and it is important to note that although the post-IVM read count was greater than 1.5x the sample mean, the pre-treatment read count was reduced (0.25x sample mean). The region was located within the first intron of a 10-exon gene (Table 6.21).

CNV regions identified on Farm 3 were frequently larger than those on Farm 2, ranging from just under 2 kb to almost 15 kb wide. Four CNV regions had the majority of reads aligning with  $\text{MAPQ} > 20$  (Table 6.21). One CNV region located on Tci2 contig 209 appeared to cover most of *g18981*, a predicted homologue of *HCON\_00078870*. Two other regions were within introns of genes on contigs 382 (*g24918*) and 755 (*g34397*). *C. elegans* homologues of these genes were predicted to be *nep-1* and *drh-3*. Two-stage RNAi of *nep-1* in *dgk-1* mutants confers resistance to paralysis in the presence of the acetylcholinesterase inhibitor, aldicarb (Sieburth et al., 2005). This *nep-1* homologue, *g24918*, was previously identified within a genetically differentiated region common to both farms (Table 6.14). The fourth CNV region at the 5' end of Tci2 contig 2721, was approximately 5 kb in width and separated from the rest of the contig by a long string of NNNs. No genes were present within the region.

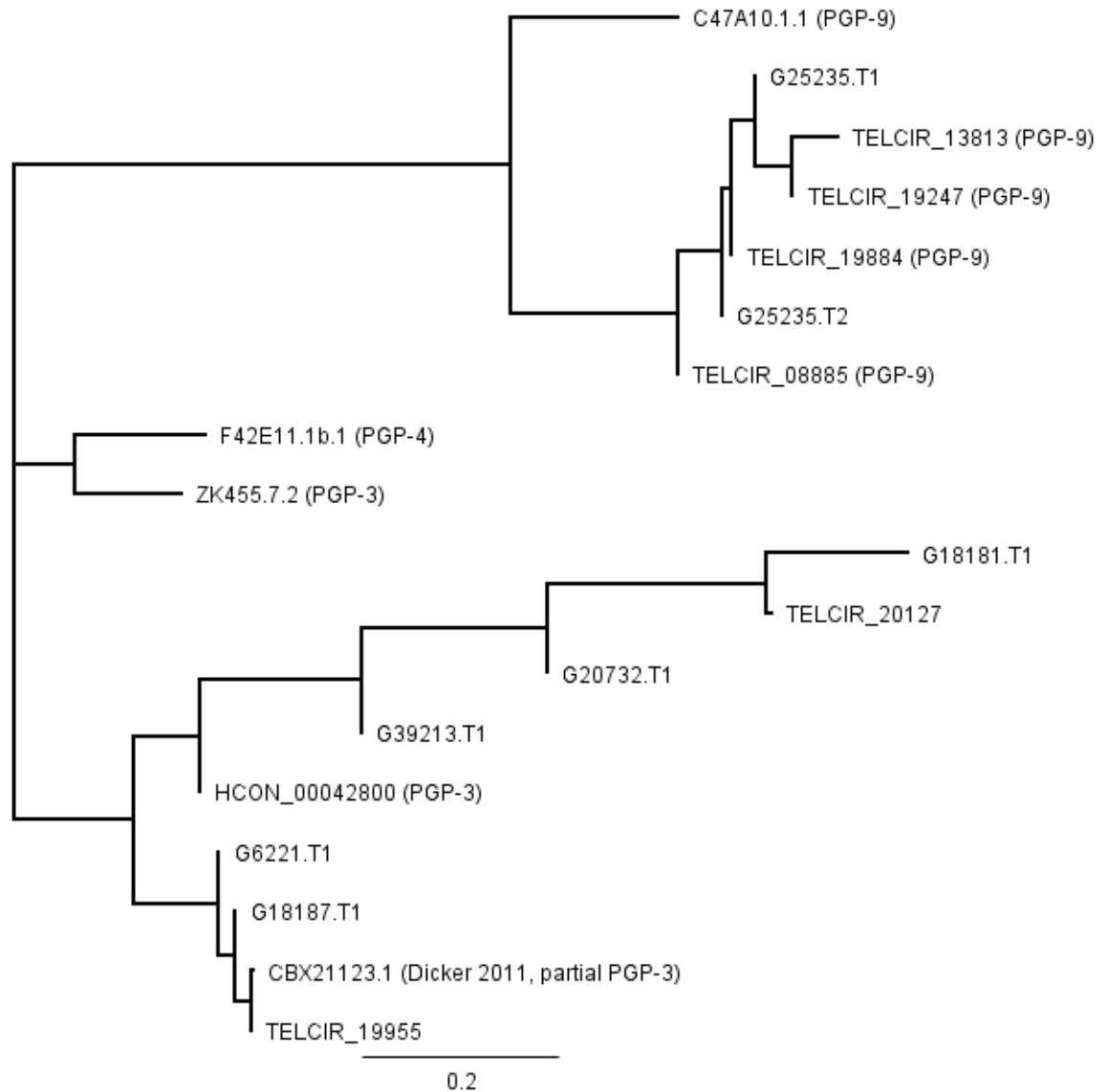
**Table 6.21: Genes and their predicted homologues/equivalent PRJNA72569 genes within or containing a region of altered copy number by read count.**

Tci2 contig CNV window <sup>1</sup>	Average log <sub>2</sub> read count ratio (Post:Pre) <sup>4</sup>	Tci2 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	Gene description	Actual CNV length	Notes
209:532501-537500 <sup>2</sup>	1.28-1.33	<i>g18981</i>	NA	<i>HCON_00078870</i>	NA	SET domain	5185 bp	Almost covers Tci2 gene annotated.
382:262501-267500 <sup>2</sup>	1.24-1.30	<i>g24918</i>	<i>TELCIR_03613</i>	NA	<i>ZK20.6 (nep-1)</i>	Endothelin converting enzyme	12746 bp	Within an intron (3/4) <sup>5</sup>
755:307501-312500 <sup>2</sup>	1.19-1.44	<i>g34397</i>	<i>TELCIR_18240</i>	<i>HCON_00009250</i>	<i>D2005.5 (drh-3)</i>	DExD/H-box helicase 58	3808 bp	Within an intron (7/14) <sup>5</sup>
755:125001-132500 <sup>3</sup>	3.63	<i>g34388</i>	<i>TELCIR_06296</i>	<i>HCON_00127120</i>	<i>ZK185.5</i>	Cation transmembrane transporter activity	98 bp	Within an intron (1/9) <sup>5</sup>

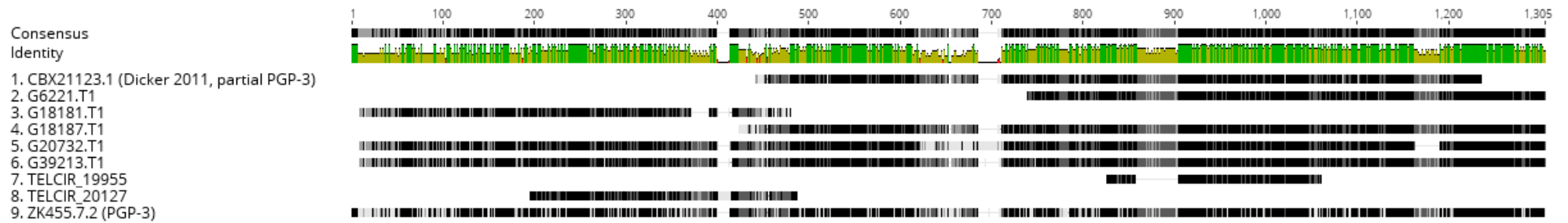
Comments are provided based on visualisation of aligned BAM files in Artemis. <sup>1</sup>Contig:nucleotide range. <sup>2</sup>Region identified on Farm 3, all pairwise comparisons. <sup>3</sup>Region identified on Farm 2. <sup>4</sup>The range is provided for Farm 3 regions. <sup>5</sup>Denotes (intron containing CNV / total number of introns) TE = Transposable element.

*Tci-pgp-9* (*g25235*) was predicted to be single-copy in the Tci2 genome following identification of Tci2 equivalent genes of the four *pgp-9* homologues identified in the PRJNA72569 genome by Choi et al. (2017) (Figure 6.2, Appendix 15). *g25235* was located on Tci2 contig 391:57466-77288. This contig was not present in any of the CNV-Seq results previously filtered. To understand why *pgp-9* was not present in the filtered CNV results, data for this region were extracted, including regions 391:57350-77300. Differences between pre- and post-treatment samples appeared minimal ( $\log_2(\text{read count ratio})$  0.15 to -0.21), and for most windows, read count ratios were non-significant. Nevertheless, the mean read count over *g25235* was higher than the mean read count for each sample, which might indicate increased copy number of *pgp-9* within the population as a whole relative to MTci2, or the presence of falsely collapsed paralogues within the genome.

Another P-glycoprotein, *pgp-3*, had been previously identified by  $F_{ST}$  analysis of both Farm 2 and Farm 3 data. In addition, changes in expression of *Tci-pgp-3* had been described in a study comparing a sensitive and a multi-drug resistant isolate of *T. circumcincta* (Dicker et al., 2011b). Nevertheless, *pgp-3* was not present in the filtered CNV data. Read CNV data were extracted for each of *g6221*, *g18187*, *g20732* and *g39213*, identified as full or partial homologues of *pgp-3* (Figure 6.2 and Figure 6.3, Appendix 15). Overall there was considerable variation across the length of the genes as to whether the  $\log_2(\text{read count ratio})$  was increased or reduced. Values tended to vary between -0.5 and 0.5 for all four genes examined indicating only small differences and most were not deemed to be significant. In comparison to *pgp-9*, the mean read count for each gene was either the same or lower than the mean read count for each sample, which likely reflects the multiple *pgp-3* haplotypes remaining in the Tci2 draft genome, in contrast to *pgp-9*, which was assembled as a single gene. In conclusion, there is no evidence of selection for CNV of *pgp-9* or *pgp-3* with IVM treatment in this study.



**Figure 6.2: A Neighbour-Joining tree of P-glycoprotein amino acid sequences. Using Geneious® v9.1.8, the amino acid sequences of the top two PRJNA72569 proteins (TELCIR\_19955 and TELCIR\_20127) identified as partial homologues of CEL-PGP-3, were compared with those identified as PGP-3 within the Tci2 genome (G6221.T1, G18187.T1, G39213.T1, G20732.T1 and G18181.T1). A partial sequence of TCI-PGP-3 by Dicker et al. (2011b) was also included for comparison. CEL-PGP-4 was included as Dicker et al. suggested that the partial sequence had similarity to both CEL-PGP-3 and CEL-PGP-4. Both PGP-9 isoforms identified in Tci2 are shown (G25235.T1 and .T2), with the PRJNA72569 PGP-9 proteins (TELCIR\_13812, TELCIR\_19247, TELCIR\_19884 and TELCIR\_08885) and the *Caenorhabditis elegans* PGP-9. Genetic distance model: Jukes-Cantor, no outgroup was used.**



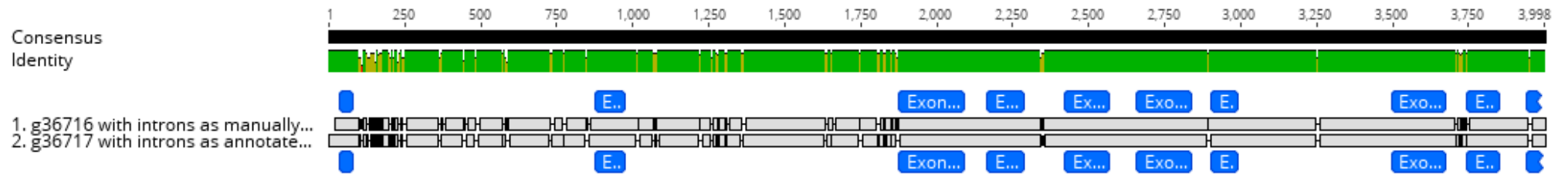
**Figure 6.3: Alignment of amino acid sequences of Tci2 gene copies of PGP-3.** Using Geneious® v9.1.8, the amino acid sequences of CEL-PGP-3, the partial PGP-3 sequence identified by Dicker et al. (2011b), and the two identified PRJNA72569 partial proteins predicted to be partial homologues of PGP-3 were compared to the five putative Tci2 PGP-3 proteins. Each amino acid sequence is shown in black where there is full consensus, grey where there is not consensus. Gaps in alignment are indicated. The consensus sequence is at the top. Sequence similarity is indicated by a coloured bar plot between the consensus and sample sequences. Where there is complete agreement between sequences this is shown in green, with reduced height and varying colour from yellow to red for decreased consensus.

### 6.3.7 Investigation of multiple gene copies in the Tci2 genome

During the course of the analysis it became apparent that the Tci2 genome contained multiple copies of genes that were annotated as single copy genes in the PRJNA72569 genome. During alignment of reads to a genome, it is possible that reads are falsely split between different gene copies, if these are haplotypes rather than paralogues. This affects the overall coverage depth and allele frequencies and can generate false positive or negative results.

To compare coverage of genes identified in the analysis to determine whether they were allelic copies or paralogues, a standard,  $\beta$ -tubulin isotype-1, was chosen for comparison. The  $\beta$ -tubulin isotype-1 gene is expected to be single copy, and has only a single copy, *TELCIR\_01271*, in the PRJNA72569 genome. However, in the Tci2 genome two gene copies of  $\beta$ -tubulin isotype-1 were identified. These were adjacent on the same Tci2 contig, separated by a short region. The read coverage over each of these genes was no different to the mean read coverage over the genome. This was concerning, as it could indicate that many genes are duplicated, with read coverage split. Although there was no difference in the amino acid sequences, the gene sequences had only 97.7% pairwise identity, primarily differing within the first two introns (Figure 6.4).

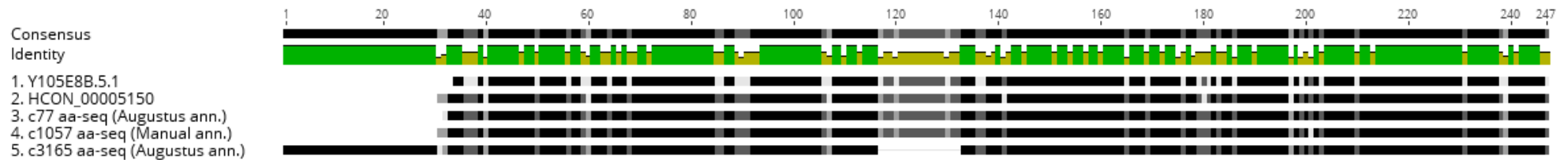




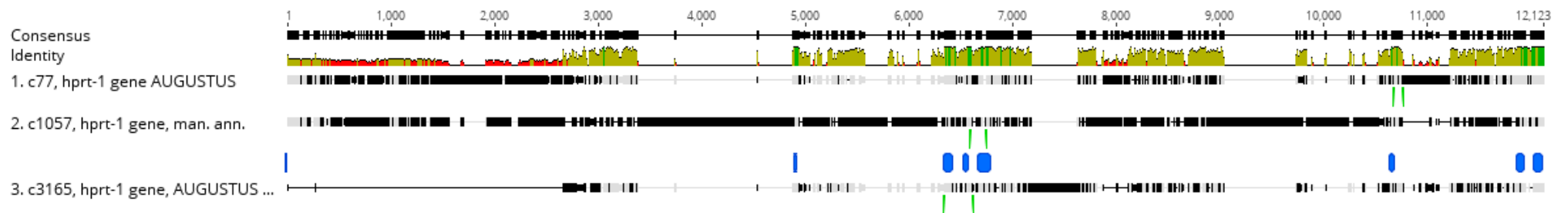
**Figure 6.4:** Pairwise alignment of the two copies of Tci2  $\beta$ -tubulin isotype-1 genomic sequence. Nucleotide sequences from the first to last exon, including introns, of the manually annotated gene copies were aligned using the 'geneious alignment' tool in Geneious<sup>®</sup> v9.1.8. Exons are indicated by blue annotations. Each sequence is indicated on a separate line, with a consensus sequence displayed at the top. Sequence similarity is indicated by a coloured bar plot between the consensus and sample sequences. Where there is complete agreement between sequences this is shown in green, with reduced height and varying colour from yellow to red for decreased consensus. Differences between sample sequences are highlighted in black, with gaps indicated (i.e. indels relative to another sequence). Regions with no difference to another sequence (or no overlapping sequence) are shown in grey.

A gene close to *TELCIR\_01271*, a homologue of *Cel-hprt-1*, had three copies within the Tci2 genome, whereas within the PRJNA72569 genome it only had one copy, with a further two partial copies of two exons each. This gene is frequently used as a normalising gene for qPCR (Stephens et al., 2011; Żyżyńska-Granica and Koziak, 2012). The Tci2 *hprt-1* copies were similar over the exonic regions (96.4% pairwise identity), with similar translated nucleotide sequences (Figure 6.5) but were very different when the full-length gene sequences were compared (42.2% pairwise identity, Figure 6.6). A simple PCR was designed to experimentally assess the number of gene copies of *hprt-1* in 95 individual Tci2 L3. Primers were designed to amplify only one of each of the Tci2 *hprt-1* gene copies as described in 2.4.5. All primer sets produced amplified products. Nevertheless, a maximum of two amplicons were present in each of the Tci2 L3 tested, as expected for a single gene in a diploid organism.

It was decided to align reads to the PRJNA72569 genome to compare results with the Tci2 analysis. Due to the more fragmented nature of this genome, and because of time constraints, a full analysis of the data was not performed. The mean read depth over *TELCIR\_01271* was twice that of the Tci2  $\beta$ -tubulin isotype-1 gene copies (~33-40x to ~73-80x). Nevertheless, the genome wide mean depth of coverage was similar for both genomes, suggesting that each had the potential for similar proportions of artefactually duplicated genomic regions.



**Figure 6.5: Alignment of HPRT-1 with orthologues.** The amino acid sequences of HPRT-1 were obtained from WormBase ParaSite for each of *Caenorhabditis elegans* (Y105E8B.5.1) and *Haemonchus contortus* (HCON\_00005150) and aligned in Geneious® v9.1.8 with the amino acid sequences as annotated by AUGUSTUS (BRAKERv2.0) of the three Tci2 HPRT-1 copies (Tci2 contigs 77, 1057 and 3165). Each sequence is indicated on a separate line, with a consensus sequence displayed at the top. Sequence similarity is indicated by a coloured bar plot between the consensus and sample sequences. Where there is complete agreement between sequences this is shown in green, with reduced height and varying colour from yellow to red for decreased consensus. Differences between sample sequences are highlighted in grey within the sequence, while regions with no difference to another sequence (or no overlapping sequence) are shown in black. Deletions are indicated as gaps in the sequence.



**Figure 6.6: Three Tci2 gene copies of *hprt-1* aligned.** The genes were extracted from the Tci2 contigs and aligned in Geneious® v9.1.8. Exons are annotated in blue. Primer positions are indicated by lime green vertical lines. Each sequence is shown on a separate line, with a consensus sequence displayed at the top. Sequence similarity is indicated by a coloured bar plot between the consensus and sample sequences. Where there is complete agreement between sequences this is shown in green, with reduced height and varying colour from yellow to red for decreased consensus. Differences between sample sequences are highlighted in black within the sequence, with gaps indicated (i.e. indels relative to another sequence). Regions with no difference to another sequence (or no overlapping sequence) are shown in grey.

### 6.3.8 Comparison of genes identified within each genome alignment analysis

Results obtained following alignment to each of the *T. circumcincta* genomes were compared. Genes within regions of genetic differentiation by  $F_{ST}$  analysis for each alignment were extracted and compared. Nine equivalent genes were identified in both Farm 2 alignment analyses (Table 6.22). Two of these genes in the Farm 2 comparison were LGICs; *lgc-39* and *lgc-42*. The gene *TELCIR\_05875*, identified as the equivalent gene of both Tci2 *g2278* and *g33217*, was predicted to be a homologue of *HCON\_00155250*, an oxidoreductase (Table 6.22). This gene is located centrally within the Chromosome V locus of genetic differentiation, which appears to be under selection by IVM treatment in *H. contortus* (Laing, *pers comm*). A single gene was identified in Farm 3 data, *gcy-22*, a receptor-type guanylate cyclase (Table 6.23).

Comparison of the PRJNA72569 alignment analysis performed in this study with the NZ study by Choi et al. (2017) revealed that no genes within regions of differentiation were identified in both studies. No ddRAD-Seq SNPs identified by Choi et al. (2017) were found to reside within the high  $F_{ST}$  10 kb windows in this study. However, one PRJNA72569 contig contained a ddRAD-Seq SNP identified by Choi et al. (2017) and also contained a high  $F_{ST}$  10 kb window in the Farm 2 data. The ddRAD-Seq SNP was PRJNA72569 contig 93:73526 and the 10 kb window in this study was located at 93:130000 - 140000 bp. No gene was annotated within this window. Two contigs were identified which contained a differentiated region in the Farm 3 PRJNA72569 alignment data and a SNP in the ddRAD-Seq data of Choi et al. (2017). The ddRAD-Seq SNPs were PRJNA72569 contig 1:1171769, with a 10 kb window located at 1:210000 - 220000 bp, containing a TE, and PRJNA72569 contig 29:339261, with a 10 kb window located at 29:500000 - 510000 bp. The latter window was covered by a 17-exon gene, of unknown function.

**Table 6.22: Consensus regions potentially under ivermectin selection identified on Farm 2 following alignment to both *Teladorsagia circumcincta* reference genomes.**

Tci2 genomic location <sup>1</sup>	Tci2 transcript	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 genomic location <sup>1</sup>	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	Gene description
5446:0-10000	<i>g10689.t1</i>	TELCIR_16059	7206:233-26210	NA	NA	Rho guanyl-nucleotide exchange factor activity
6704:0-10000	<i>g12034.t1</i>	TELCIR_16059	7206:233-26210	HCON_00054950 <sup>2</sup>	K07D4.7 ( <i>ephx-1</i> ) <sup>2</sup>	Ephexin (Eph-interacting GEF) homolog
6871:0-10000	<i>g12159.t1</i>	TELCIR_13217	4303:7822-25124	HCON_00101310	F36H1.6 ( <i>alh-3</i> )	10-formyltetrahydrofolate dehydrogenase
229:0-10000	<i>g19789.t1</i> <i>g19789.t2</i>	TELCIR_15799	6890:483-18301	HCON_00089890	Y39A3B.2 ( <i>lgc-42</i> )	LGIC
30:620000-640000	<i>g2278.t1</i>	TELCIR_05875	756:104825-112625	HCON_00155250	E02C12.19	CHK kinase-like, uncharacterised oxidoreductase Dhs-27
708:120000-140000	<i>g33217.t1</i>	TELCIR_05875	756:104825-112625	HCON_00155250	E02C12.19	CHK kinase-like, uncharacterised oxidoreductase Dhs-27
797:130000-140000	<i>g35340.t1</i>	TELCIR_14359	5320:17976-24085	HCON_00125890	F29B9.4 ( <i>psr-1</i> )	Bifunctional arginine demethylase and lysyl-hydroxylase
928:180000-190000	<i>g37932.t1</i> <i>g37932.t2</i> <i>g37932.t3</i> <i>g37932.t4</i>	TELCIR_14226	5200:152-24827	HCON_00145000	F09G2.5 ( <i>lgc-39</i> )	LGIC
1242:150000-160000	<i>g42660.t1</i>	TELCIR_13482	4516:9020-30978	HCON_00121072	R06F6.4 ( <i>set-14</i> )	SET domain-containing protein 14

Tci2 genomic location <sup>1</sup>	Tci2 transcript	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 genomic location <sup>1</sup>	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	Gene description
1242:150000-160000	<i>g42661.t1</i>	TELCIR_13482	4516:9020-30978	HCON_00121072	R06F6.4 ( <i>set-14</i> )	SET domain-containing protein 14
76:420000-430000	<i>g5091.t1</i>	TELCIR_09340	1909:54809-65032	HCON_00142160	H19N07.1 ( <i>erfa-3</i> )	Eukaryotic Release Factor homolog
2498:80000-90000	<i>g52925.t1</i>	TELCIR_11587	3104:34-34009	HCON_00144870	T01D3.1	Calcium ion binding. Motor neuron axon guidance

Reads were aligned to both available *Teladorsagia circumcincta* reference genomes. Tci2 genes identified within regions of genetic differentiation were BLASTP searched for equivalent genes in the PRJNA72569 genome. Those identified were compared to the list of PRJNA72569 genes within regions of genetic differentiation in the PRJNA72569 alignment analysis. <sup>1</sup>Contig:nucleotide range. <sup>2</sup>Note, these are the predicted orthologues of *TELCIR\_16059*. However, they were not identified during BLASTP search using the WSI *farm\_blast* for G10689.T1.

**Table 6.23: Consensus regions potentially under ivermectin selection identified on Farm 3 following alignment to both *Teladorsagia circumcincta* reference genomes.**

Tci2 genomic location <sup>1</sup>	Tci2 transcript	<i>Teladorsagia circumcincta</i> PRJNA72569 gene	<i>Teladorsagia circumcincta</i> PRJNA72569 genomic location <sup>1</sup>	<i>Haemonchus contortus</i> gene	<i>Caenorhabditis elegans</i> gene	Gene description
155:330000-340000	<i>g16586.t1</i>	<i>TELCIR_04718</i>	496:50000-60000	<i>HCON_00131630</i>	<i>T03D8.5 (gcy-22)</i>	Receptor-type guanylate cyclase <i>gcy-22</i>

Reads were aligned to both available reference genomes. Tci2 genes identified in the first analysis were BLASTP searched for equivalent genes in the PRJNA72569 genome. Those identified were compared to the list of PRJNA72569 genes within regions of genetic differentiation in the PRJNA72569 alignment analysis. <sup>1</sup>Contig:nucleotide range.

## 6.4 Discussion

Many studies have explored IVM resistance in laboratory isolates and field populations of different nematode species, with no clear consensus to date. Of the genes identified, most have been experimentally examined in either *C. elegans* or *H. contortus* with only a few also investigated in *T. circumcincta*. The aim of this study was to identify potential markers of IVM resistance using WGS. Following bioinformatic analysis, over 600 genes were identified within regions of genetic differentiation on both farms. Some regions were the same in each farm analysis, however most were different. Genes potentially related to IVM resistant phenotypes were identified and are discussed below. Note that a ranking is indicated for certain genes, relating to the  $F_{ST}$  value of the 10 kb window containing the gene. Although this provides one opportunity to determine how important a gene may be for IVM resistance, it does not necessarily correspond directly. For example, in *P. falciparum* populations, the *kelch* gene has undergone a soft sweep, with multiple alleles containing non-synonymous SNPs selected by artemisinin treatment, reducing  $F_{ST}$  values relative to nearby genes (Cheeseman et al., 2015). All ranking data can be found within Appendices 6 and 7.

### 6.4.1 Neuronal genes

Key IVM candidate genes include the GluCl receptors (Dent et al., 1997; Dent et al., 2000). A single GluCl gene was identified in this study, a predicted homologue of *Cel-glc-3* (rank 172 and 200/576, Farm 2 analysis). A related gene, *glc-4*, was recently identified when comparing IVM sensitive and resistant *H. contortus* individuals (Sallé et al., 2019). *Cel-glc-3* was originally identified as an IVM sensitive GluCl, when it was shown to form a homomeric receptor in the *Xenopus* oocyte expression system (Horoszok et al., 2001). A response to IVM was elicited at IVM concentrations of  $> 0.1 \mu\text{M}$ , but not at  $0.01 \mu\text{M}$ . In *C. elegans* this gene appears to be involved in regulating both locomotion (Cook et al., 2006) and searching behaviour (Chalasani et al., 2007), with both deletion mutants or increased expression of *glc-3* affecting this behaviour. It is expressed in the AIY interneurons, part of the pathway controlling normal searching behaviour and duration (Wenick and Hobert, 2004; Chalasani et al., 2007). In *H. contortus*, a



small increase in expression of *glc-3* was reported in a triple resistant (BZ, LEV, IVM) isolate (UGA/2004), compared to a drug susceptible isolate (US/HcS) (Williamson et al., 2011). *glc-3* may therefore be of interest in IVM resistance, but its role is difficult to interpret, as it has not been identified consistently in other studies and was not identified in the Farm 3 analysis. Other Tci2 genes potentially involved in chemosensory pathways were identified, including *acr-15* (Sellings et al., 2013), *deg-3/des-2* (Yassin et al., 2001), *snf-3* (Peden et al., 2013), and *R05F9.7* in the Farm 2 analysis and *snf-10* (Yemini et al., 2013) in the Farm 3 analysis.

Other LGIC have been associated with IVM resistance including GABA receptors in *H. contortus* (Blackhall et al., 2003) and *lgc-54*, for which the ligand is unknown in *T. circumcincta* (Choi et al., 2017). Analysis of Farm 2 data identified two LGICs, predicted to be *lgc-39* (rank 322/576) and *lgc-42* (rank 285/576). Importantly, these were identified following alignment of reads to either *T. circumcincta* genome. SNPs above the Farm 2 ddRAD-Seq  $F_{ST}$  threshold were identified ~80 kb distant to *lgc-39* ( $F_{ST}$  0.1, odds ratio 0.1), and ~320 kb distant to *lgc-42* ( $F_{ST}$  0.1, odds ratio 0.1). If these contigs have been correctly assembled, this may suggest a wider region of selection surrounding these genes. Furthermore, increased expression of *Hco-lgc-42* in IVM resistant male *H. contortus* (2.14 fold) has been observed compared to untreated individuals (Laing, *pers comm*). Little information is available regarding these two LGICs in the literature, however both are predicted to be GABA or glycine receptors. Therefore, although these genes may be correlated with IVM resistance on Farm 2, their role in IVM sensitivity is difficult to predict.

A third LGIC identified as part of the Farm 2 analysis was *lgc-27* (rank 66/576). This also had a ddRAD-Seq SNP identified ~300 kb distant ( $F_{ST}$  0.18, odds ratio 13.8). Comparison with Choi's study identified two SNPs on the same PRJNA72569 contig as the equivalent gene (Choi et al., 2017), however *lgc-27* was not identified when Farm 2 reads were aligned to the PRJNA72569 genome. *Cel-lgc-27* is not well characterised. RNAi of *lgc-27* extended the lifespan of mutant dauer larvae, which would usually be expected to die after 10-12 days following rapid hydrolysis of lipid stores (Xie and Roy, 2012). This gene may play a role in prolonging survival during periods of starvation, which might occur

following exposure to IVM. A fourth Tci2 gene was predicted to be *Cel-lgc-41* (rank 549/576). However, this gene is more likely to be a homologue of *Hco-unc-49*. In *C. elegans*, the protein UNC-49 is a GABA receptor subunit for which IVM was previously identified to be a ligand (Hernando and Bouzat, 2014). Expressed in L1 muscle cells, IVM appeared capable of antagonising the effect of both GABA and piperazine on muscle, apparently irreversibly, but did not act as an agonist itself. Although this gene has a comparatively low ranking by  $F_{ST}$ , it does have a clearer association with IVM than many others and may be of interest for further study. No LGICs were identified in the Farm 3 analysis.

IVM resistance in *C. elegans* has been demonstrated with mutations in three genes affecting pharyngeal pumping (Dent et al., 2000). In parasites, two *in vitro* tests are available to assess IVM phenotype which rely upon the action of IVM to induce pharyngeal paralysis and starvation (Gill et al., 1995; Alvarez-Sanchez et al., 2005b). Two neuronal genes related to pharyngeal pumping were identified, which may be involved in IVM resistance. *gar-3*, a muscarinic AChR, was identified in the Farm 2 analysis (rank 55/576) (Hwang et al., 1999; Park et al., 2003). *gar-3* appears to play an important role in enabling sufficient time for filling of the pharyngeal terminal bulb with food (Steger and Avery, 2004). A gene predicted to be a homologue of the neprilysin *Cel-nep-1* was identified in both farm analyses, although the exact homology is uncertain. This gene was within the highest ranked region by  $F_{ST}$  in the Farm 3 analysis. Neprilysins are transmembrane zinc metalloendopeptidases which catalyse hydrolysis of peptides expressed on the extracellular plasma membrane, turning off signalling events (Turner et al., 2001). In *C. elegans*, 26 neprilysin genes are listed on WormBase and there are several neprilysin genes with differing roles present in humans (Turner et al., 2001). Expression of *Cel-nep-1* was found within the pharyngeal cells in all life stages, and also in a single head neuron in adults (Spanier et al., 2005). In *Cel-nep-1* knockout mutants, a nine-fold reduction in pharyngeal basal activity was identified by electropharyngeogram compared with wildtype worms. Evidence for a link with IVM comes from studies on filarial worms, where following exposure to 100 nM IVM *in vitro*, a reduction in expression of *Bma-nep-1* of  $-3.92 \log_2$  fold was apparent in adult female *Brugia malayi* compared to unexposed worms (Ballesteros et al., 2016). Choi et al. (2017) detected an increase in expression of *TELCIR\_03613* ( $\log_2$  ratio 3.22) in

their triple resistant strain compared to their  $S_{\text{inbred}}$  strain. *TELCIR\_03613* is the predicted equivalent gene of the Tci2 neprilysin identified in both farm analyses. In contrast, in *H. contortus* RNA-Seq studies a reduction in expression of a neprilysin, *Hco-nep-22* (-0.26 to -0.68 fold change), was reported in IVM treated adults (Laing, *pers comm*). A CNV within an intron of the Tci2 *nep-1* homologue was identified in the Farm 3 analysis. In total, five neprilysins were identified in the Farm 2 ddRAD-Seq and Pool-Seq analyses. The mode of action of IVM on these genes is unclear, but they may be of interest for further study.

*lev-9*, a secreted protein essential in *C. elegans* for clustering of AChRs at synapses, and which is expressed in body wall muscle cells (Gendrel et al., 2009), was one of only three ‘neuronal’ genes identified by GO terms in the Farm 3 analysis. It is possible that alteration of this protein may affect expression or clustering of AChRs and the worm’s response to anthelmintics. However, AChRs are associated primarily with LEV resistance (Fauvin et al., 2010; Boulin et al., 2011; Barrère et al., 2014; Blanchard et al., 2018) and MPTL resistance (Kaminsky et al., 2008; Rufener et al., 2009b; Turnbull et al., 2019). AChR subunit genes associated with LEV resistance and MPTL resistance are in distinct groups (reviewed by Holden-Dye et al. (2013)). Three of the four AChR subunit genes found in the Farm 2 analysis (*acr-17*, *deg-3* and *des-2*) belong to the group associated with MPTL resistance. Both *deg-3* and *des-2* are expressed neuronally, within the PVD neuron and chemosensory neurons of the head region (IL2 and FLP cells), and within cell bodies of the tail region (Treinin et al., 1998; Yassin et al., 2001). Mutations in *deg-3* in *C. elegans* can result in neuronal degeneration (Treinin and Chalfie, 1995) and uncoordinated movement (Chalfie and Sulston, 1981; Chalfie and Wolinsky, 1990). The fourth AChR, *acr-15*, is involved in chemosensory responses to nicotine in *C. elegans* (Sellings et al., 2013), and also affects body size and locomotion (Yemini et al., 2013). Both *deg-3* (Treinin and Chalfie, 1995) and *acr-15* (Sellings et al., 2013) are potential homologues of a mammalian  $\alpha 7$  nAChR subunit. IVM has been shown to bind these subunits as an allosteric modulator, enhancing the binding of the ligand choline when functionally expressed in *X. laevis* oocytes (Krause et al., 1998; Collins and Millar, 2010). Mutations in the transmembrane domain of *deg-3* limited the effect of IVM, while other mutations facilitated inhibition by IVM of the receptor (Collins and Millar, 2010). Although IVM resistance has not been

associated with AChRs in nematodes, a recent study identified that abamectin was capable of antagonising nicotinic AChRs from *Oesophagostomum dentatum* when expressed in *Xenopus* oocytes (Abongwa et al., 2016). Interestingly, the response varied with drug concentration. At low doses, abamectin was antagonistic, however at higher doses ( $> 0.3 \mu\text{M}$ ), this antagonism reduced. In addition, IVM induced profound inhibition upon the frequency of channel opening of levamisole AChRs in *C. elegans* (Hernando and Bouzat, 2014).

IVM is also known to effect reproduction (Duke et al., 1991; Scott et al., 1991; Sutherland et al., 2003a). *gar-3* is involved in protraction of the *C. elegans* male spicule protractor muscles and male mutants were shown to have reduced persistence in vulval penetrating behaviour compared with wild type worms (Liu et al., 2007). A second gene, *lov-1*, identified within the Farm 2 analysis was also predicted by GO term annotation to affect male mating behaviour. In the Farm 3 analysis, a gene predicted to be *snf-10* (rank 37 and 59/91) may be involved in reproduction. Sallé et al. (2019) also identified a *snf* gene in their GWAS. Sodium:neurotransmitter symporters are proteins which co-transport sodium with another molecule, for example GABA, recycling neurotransmitters and terminating synaptic transmission (Bröer and Gether, 2012). *Cel-snf-10* appears important for sperm activation within the male (Fenker et al., 2014). Although mating between a male and a hermaphrodite *C. elegans* did not show immediate differences in numbers of progeny, over time the *snf-10* mutants showed reduced numbers of progeny (Fenker et al., 2014). Two other sodium:neurotransmitter symporters were identified, both in the Farm 2 analysis. *snf-3* (rank 59/576) is a betaine transporter (Peden et al., 2013). *snf-3* may have an interaction with *acr-23*, the AChR subunit that confers MPTL resistance in *C. elegans* (Kaminsky et al., 2008), as betaine was toxic to *snf-3* mutants but not to *snf-3;acr-23* double mutants (Peden et al., 2013). *C. elegans snf-3* mutants have locomotory defects and retarded larval development (Peden et al., 2013). *snf-3* is expressed in the excretory canal and tail hypodermal cells, the epidermis and vulval epithelial cells, and in several neurons, including sensory neurons of the head (Peden et al., 2013). The second *snf* gene in the Farm 2 data (rank 239/576) had an *H. contortus* homologue predicted but no *C. elegans* homologue was found. This gene family may be involved in IVM resistance, however the Tci2 genes and their predicted identity differed

between farms and there is no evidence of IVM interacting with these symporters in the literature.

### 6.4.2 Metabolic and transporter genes

IVM is known to induce starvation via cessation of pharyngeal pumping (Gill et al., 1995; Alvarez-Sanchez et al., 2005b). Survival may be dependent upon lipid stores and metabolism of these (Laing et al., 2012; Smus et al., 2017). Multiple genes were predicted to be involved in lipid metabolism in the present study. These included *acs-7*, a fatty acid CoA synthetase, a gene in the same class as *acs-16* identified by Doyle et al. (2017). Genes in this family were reported to be upregulated by Laing et al. (2012) following exposure of an IVM resistant *C. elegans* strain (DA1316) to IVM.

In other organisms, associations between drug resistance and ubiquitination have been identified. For example, in *P. falciparum*, artemisinin induces ubiquitination and resistant parasites have decreased ubiquitination and delayed onset of cell death (Dogovski et al., 2015). IVM has been shown to induce ubiquitination of P21-activated kinase 1 (PAK1), resulting in autophagy of breast cancer cells (Dou et al., 2016). On both farms, homologues of the *C. elegans* gene *usp-50* were identified in the loci under IVM selection. *usp-50* is a ubiquitin specific protease expressed in intestinal cells (Huynh et al., 2016). RNAi of *usp-50* in *C. elegans* embryos increased ubiquitination of the multi-drug resistance MRP-4 protein 7.7 fold (Huynh et al., 2016). P-gp half-life was reported to be increased following prevention of P-gp ubiquitination in a mammalian cancer cell line (Ravindranath et al., 2015). It is therefore possible that these Tci2 homologues of *usp-50* may be selected as part of a mechanism of survival, related to detoxification or reduced uptake, in the presence of IVM. Other genes also predicted to be within ubiquitination pathways by their GO terms included *F10G7.10*, *ulp-4* and *HCON\_00053760* in the Farm 2 analysis.

*H. contortus* and *C. elegans* produce glycosidated metabolites of BZ anthelmintics when exposed *in vitro* (Cvilink et al., 2008; Laing et al., 2010). Increased quantities of these metabolites were reported following selection of an IVM resistant strain (ISE-S) from the parental isolate, MHco3(ISE) (VokřÁL et

al., 2012). In a bacterial system (*Saccharopolyspora erythraea*), it has been shown that both the mono- and disaccharides of IVM accept a glycosyl group, a reaction catalysed by a constitutively expressed glycosyltransferase (Schulman et al., 1993). In the present study a predicted homologue of *Cel-ugt-49*, a UDP-glycosyltransferase, was identified in the Farm 3 analysis (rank 35 and 74/91). However, as no IVM glycoside metabolites were identified in an experiment involving *H. contortus* and *C. elegans* (Laing, 2010), the relevance is unclear. It may be that *ugt-49* is related to survival during starvation, as downregulation of genes in this class in IVM-exposed *C. elegans* DA1316 was reported previously (Laing et al., 2012).

Predicted homologues of the enzyme CEL-PYR-1 were identified in both farm analyses (rank 195/576 and 81/91). *Cel-pyr-1* mutants have a shorter (~17%) and thicker pharyngeal isthmus compared with wild type worms (Franks et al., 2006), and disorganised actin filament structure is apparent. If parasites predominantly uptake or ingest IVM through pharyngeal structures, changes to this may play a role in IVM phenotype.

The ability to rapidly excrete IVM may be a survival advantage *in vivo* in the host or in faeces. Several studies have associated P-gps with anthelmintic resistant phenotypes (Blackhall et al., 2008; Bartley et al., 2009; Sarai et al., 2013). In this study only *pgp-3* was identified as differentiated between pre- and post-IVM samples. *Tci-pgp-3* was previously identified to be significantly upregulated in eggs (3.90 fold,  $p < 0.05$ ) of the multi-drug resistant MTci5 isolate, and significantly downregulated in L4 (0.68 fold,  $p < 0.01$ ) compared with the MTci2 isolate, as determined by qPCR (Dicker et al., 2011b). In contrast, *pgp-9*, which is more strongly associated with anthelmintic resistance in *T. circumcincta* (Dicker et al., 2011b; Turnbull et al., 2018), was not identified, either by  $F_{ST}$  or by CNV, unlike in Choi's GWAS (Choi et al., 2017). This may suggest that *pgp-9* is already increased in copy number within the field populations examined in this study, and is more important as a multi-drug resistance mechanism rather than an IVM specific mechanism. Other transporters were also noted within the results. These included *abt-2* and *kcc-1* in the Farm 2 analysis, and members of the peptide transporter family in both farm analyses.

Two oxidoreductases were identified, one of which, identified only in the Farm 2 analysis, was predicted to be a homologue of the *H. contortus* gene *HCON\_00155250*. This gene is located within the Chromosome V peak, and a reduction in expression in IVM treated *H. contortus* adults (-0.73 and -1.09 log<sub>2</sub> fold, in males and females respectively) was reported (Laing, *pers comm*). Importantly, this gene was also identified in the Farm 2 analysis following alignment of reads to the PRJNA72569 genome. The association with IVM resistance of these genes is unclear, however increased expression of genes related to ‘oxidoreductase activity’ and reduction in expression of genes related to ‘oxidoreductase’ were reported by Laing et al. (2012), following exposure of *C. elegans* DA1316 to IVM.

### 6.4.3 Regulatory genes and transposable elements

Multiple genes were predicted to have regulatory function in both farm analyses, including genes of the ‘abnormal cell lineage’ class. One of these, identified in the Farm 2 analysis, is predicted to be a homologue of *Cel-lin-44*, part of the WNT signalling pathway that can be important during development. This pathway has been reported to be inhibited by IVM in a study of colorectal cancer (Melotti et al., 2014). Other regulatory genes included nuclear hormone receptors. A nuclear hormone receptor, *nhr-8*, was recently suggested to play a role in IVM tolerance in *C. elegans* (Ménez et al., 2019). Two genes in the Farm 2 analysis, predicted to be homologues of *Cel-mab-10*, a transcriptional co-factor (Harris and Horvitz, 2011), were identified (rank 4 and 219/576 and 351/576). Of these, one was also identified in the Farm 3 analysis. The PRJNA72569 equivalent of this gene contained a ddRAD-Seq SNP in the NZ study by Choi et al. (2017). Originally identified in male mutants incapable of mating, which had slightly swollen bursas (Hodgkin, 1983), *Cel-mab-10* has since been characterised as important in regulation of moulting in *C. elegans*. Mutants synthesise an extra adult cuticle, which males subsequently moult, but which hermaphrodites often fail to shed (Harris and Horvitz, 2011). Expressed in pharyngeal cells, vulval precursor cells, seam cells, gonads and hypodermal nuclei (Harris and Horvitz, 2011), the link with IVM resistance is not clear. Potentially, alterations in the cuticle might enhance the ability of worms to survive IVM treatment due to restricting IVM entering the worm (e.g. failure to shed a cuticle might impair

uptake via ingestion/amphids/phasmids). Alternatively, it may be important in regulatory pathways associated with IVM resistance.

Several genetically differentiated regions identified in both farm analyses contained TEs. These are DNA sequences with the inherent ability to self-replicate, altering their genomic position and inducing structural changes within a genome (reviewed by Bourque et al. (2018)). Of the total number of proteins identified to be present in both farm datasets approximately a third were predicted to be TEs. These results may reflect true selection, or confounding appearance of selection due to misalignment of reads over the Tci2 TE, without necessarily reflecting the genomic position of the TE(s) within the field populations (Laing et al., 2011). Also, some TEs may coincidentally appear within a region of genetic differentiation due to a general abundance within the genome, without being the cause of selection in the region. TEs were previously demonstrated to play a role in insecticide resistance in *Drosophila melanogaster* (reviewed by Rostant et al. (2012)). Here, the presence of strong selection around a retrotransposon, *Accord*, inserted into the 5' UTR of the *cyp6g1* gene was associated with DDT resistance (Daborn et al., 2002; Schlenke and Begun, 2004). In *Drosophila simulans*, a *Doc* long terminal repeat retrotransposon in the 3' UTR of the gene upstream of *cyp6g1* was also associated with DDT resistance (Schlenke and Begun, 2004). These insertions were shown to increase transcription of the *cyp6g1* gene (Schlenke and Begun, 2004), presumably enhancing metabolism of DDT. In contrast, an insertion of *Doc1420* TE into the second exon of a choline kinase gene *CHKov1*, changed the transcription pattern from a single transcript to two transcripts within an individual *D. melanogaster* (Aminetzach et al., 2005). The presence of this TE was associated with resistance to an organophosphate, azinphos-methyl-phosphate (Aminetzach et al., 2005). TEs are abundant within the related nematode *H. contortus*, with differences in insertion patterns evident between isolates and to a certain extent also between derived strains and their parental populations (Hoekstra et al., 2000). That some of these TEs are still active has been demonstrated by the presence of TEs within RNA-Seq data of *H. contortus* (Laing et al., 2011). TEs may be involved in IVM resistance in *T. circumcincta*, however analyses more suited to exploration of TE differences are required before conclusions could be



drawn (Kofler et al., 2016a; Bourque et al., 2018), and would require an improved reference genome.

#### 6.4.4 Multiple gene copies have potential to confound analysis

The presence of Tci2 genes with many-to-one homology with genes in *C. elegans* was noticed during the course of the analysis. These included genes that might be expected to have paralogues, such as *pgp-3*, but also genes which were expected to be single copy. If genes were truly paralogues, the presence of multiple gene copies would be real, and should be minimal in their effect on the analysis. However, if haplotypic copies of genes were retained as part of the Tci2 draft assembly, such that multiple copies were technical artefacts, the analysis could be severely impacted. Instead of reads aligning in a single stack (true alignment), reads would be split over the multiple gene copies by similarity to the haplotype, such that alleles identified for a copy of the gene would not reflect true population allele frequencies, or even sample frequencies. This would have an effect on all downstream analyses and might help to explain some of the differences noted between genome alignments in the present study.

Support for multiple gene copies was provided by the high percentage (~17%) of reads with MAPQ=0, as reported in 5.3.2. Due to the filtering performed in this study, reads with MAPQ=0 were removed. This is likely to have affected coverage over technical haplotypes. Variant sites with <20 mean depth within a sample were not included in the SnpEff analysis, and some may have been lost due to low coverage following filtering of these MAPQ=0 reads.

Investigation of read depth over  $\beta$ -tubulin isotype-1 in each genome alignment identified two copies of  $\beta$ -tubulin isotype-1 to be present. Nevertheless, this seems unlikely, as many previous studies have investigated BZ resistance using this gene, and have not reported multiple gene copies of  $\beta$ -tubulin isotype-1 in *T. circumcincta* (Skuce et al., 2010; Redman et al., 2015). A gene close to  $\beta$ -tubulin isotype-1 in the PRJNA72569 genome also appeared to be multiple copy in the Tci2 genome, and to have partial copies in the PRJNA72569 genome. This gene, *hprt-1*, is also expected to be fairly conserved. PCRs failed to identify

more than two alleles of this gene in any MTci2 individual. Thus, although the presence of multiple gene copies cannot be excluded, the PCR results do not provide evidence for more than one gene copy within the *T. circumcincta* genome, despite the low percentage identity. This highlights the dramatic degree of diversity present within a population and the difficulties in assembling a truly representative reference genome. The extent of this issue within the Tci2 draft genome is unknown, but it is concerning as it indicates that both false positive signals could be identified, while other regions under selection by IVM could be missed.

Interestingly, the gene *pgp-3* had four or five Tci2 gene copies identified, with a read count over the gene equivalent to the genome-wide mean read count or lower. In contrast, *pgp-9* had only one copy, yet had increased read count over the gene in both the pre- and post-IVM populations in all samples compared to the median read count for each sample. These results suggest two possibilities. Firstly, if Tci2 genes are correct, Tci2 has multiple paralogues of *pgp-3* but only a single copy of *pgp-9*. Dicker et al. (2011b), identified a significant increase in expression levels of *pgp-9* mRNA in all lifecycle stages in MTci5 compared to MTci2, but did not identify the same differences when studying *pgp-3*. The Tci2 copy number identified in the genome would help to partially explain these results if MTci5 had increased copy number of *pgp-9* compared to MTci2 (Turnbull et al., 2018), but similar copy number of *pgp-3*. There are four copies of *pgp-9* identified in the PRJNA72569 genome, based on a NZ isolate, and although Choi et al. (2017) suggested increased *pgp-9* copy number in the triple resistant NZ strain used in their study, they did not identify *pgp-9* by  $F_{ST}$  analysis. It may be that the results identified in these studies reflect inherent differences between *T. circumcincta* populations rather than those directly related to anthelmintic resistance. Alternatively, the *pgp-3* gene copies could be technical artefacts, as the read count over the gene is the same or less than the mean read count and *pgp-3* may or may not be important in IVM resistance. Of course, if the *pgp-3* copy number in the field samples is less than in MTci2, then reads would be split over paralogous loci. It is also possible that the single copy Tci2 *pgp-9* represents falsely collapsed paralogues, and hence has increased coverage.

## 6.5 Conclusion

Overall, this genome wide analysis has indicated several genes that may be important in IVM resistance, but with low numbers of genes identified in both farms or that are consistent with other IVM GWAS. In addition, only a few genes were identified in both genome analyses, highlighting the impact of the genome assembly. Interpretation of the results is complex and should be performed with care due to the possibility of multiple technical gene copies and the fragmented state of the genome assemblies. However, the sequencing data and analysis pipelines remain a high quality resource and will warrant re-analysis when the Tci2 genome is complete

## 7 Discussion: Identifying the genetic basis of IVM resistance proved extremely challenging

The primary aim of this PhD was to identify genetic markers of IVM resistance in field populations of *T. circumcincta*. This proved extremely challenging. Field populations sampled differed in their IVM resistance status, and mixed species infections were common, reducing the correlation between the FECRT results and the IVM efficacy against *T. circumcincta*. As a result, locating a suitable farm to use in the genome wide association studies (GWAS) took longer than anticipated. The prevalence and severity of IVM resistance on UK farms is increasing (Sargison et al., 2005; Hybu Cig Cymru, 2015) and the first farm chosen proved to have potentially ‘fixed’ IVM resistance, with no detectable efficacy of the anthelmintic against *T. circumcincta*. Other farms tested varied between comparatively high efficacy (80% reduction in FEC) and low efficacy (20% reduction in FEC), or no efficacy against *T. circumcincta*. None were identified with a moderate efficacy of 50-60%. Unlike *H. contortus*, FECs of *T. circumcincta* infections are often comparatively low (Gibson and Parfitt, 1977). It was extremely difficult to identify a field population which had sufficient numbers of *T. circumcincta* L3 both pre- and post-IVM which also displayed an appropriate IVM efficacy for a GWAS. If IVM resistance was too high, detection of IVM resistance markers by  $F_{ST}$  comparison of pre- and post-treatment L3 was thought to be unlikely. In contrast, field populations against which IVM was still comparatively effective had limited numbers of L3 available post-IVM, restricting the number of larvae available for sequencing. Eventually, two field populations were chosen: one had a high IVM efficacy (80%, Farm 2) and the other a low IVM efficacy (20%, Farm 3). These populations were considered the best available for the experiment. However, there were additional factors which made the interpretation of results more challenging. Sub-sampling of the overall *T. circumcincta* population within the sheep hosts could have occurred at any given time point. Such that, rather than sampling individuals representative of the entire population treated with IVM at each time point, a sub-population was selected for Pool-Seq for each of the pre- and post-treatment samples. This could have occurred due to changes within the parasite community and in the faecal egg output of adult worms over time and following IVM treatment (Bishop and Stear, 2000; Sutherland et al., 2003a). The ddRAD-Seq results, however,

suggest that sub-sampling has not occurred on Farm 2, and the allele frequency correlations using the Pool-Seq data also suggest that no sample has a greater degree of differentiation to any other. An ideal sample would be comprised of an entirely homogenous population, with a subset of individuals resistant only to IVM, however both farms had multi-drug resistance. Indeed, it would be impossible to locate a UK farm without some resistance to BZ (McMahon et al., 2013b; Hybu Cig Cymru, 2015). Whether individual *T. circumcincta* were dual-resistant or multi-drug resistant was not known, but was likely, especially based on the results observed on Farm 1 and by the dual product FECRTs performed on Farm 3. Several genes identified on these farms have associations with LEV (Fleming et al., 2009; Gendrel et al., 2009), MPTL (Kaminsky et al., 2009; Rufener et al., 2009b) or derquantel activity (Puttachary et al., 2013; Abongwa et al., 2016), suggestive of prior exposure to these classes of drug. Farm management practices, such as those observed on Farm 1, may preferentially select for multi-drug resistant parasites with proportionally few individuals mono-resistant, alongside susceptible parasites. This could therefore increase differences between pre- and post-treatment samples, unrelated to IVM resistance. These unrelated loci could confound interpretation, especially as IVM has a broad spectrum of activity, and multiple gene families have been previously suggested to be associated with tolerance, target action, detoxification and resistance in the literature.

Although it was expected that Farm 2 would display greater genetic differentiation (by  $F_{ST}$ ), compared with Farm 3, between the pre- and post-IVM populations, due to a greater drug efficacy in this population, this was not observed. A potential reason for this discrepancy could be increased diversity on Farm 2 due to the recent merging of two flocks, which may both have carried IVM resistant worms. This could have increased diversity of, and around, IVM resistance loci, reducing relative  $F_{ST}$ . A different farm, with similar IVM efficacy, but which did not buy in new stock each year would have been desirable for testing, but was not available.

ddRAD-Seq was trialled as a reduced representation method and has since been published as a tool to investigate IVM resistance in *T. circumcincta* (Choi et al., 2017) and validated as an experimental method in *Schistosoma japonicum*

(Shortt et al., 2017). Given the large genome size (PRJNA72569, ~700Mb), ddRAD-Seq was an economically attractive option. Unfortunately, the high diversity encountered resulted in fewer RAD markers than desired. This was likely explained by a combination of allele drop out, due to polymorphisms in the restriction enzyme cut-site (Arnold et al., 2013; Gautier et al., 2013b), and haplotypic copies of genes in the genome affecting read alignment. Nevertheless, it was useful in determining linkage decay in the genome and has provided an excellent resource for future use when the genome assemblies are improved. ddRAD-Seq may have been affected by multiple haplotypic copies of genes or regions to a greater extent than Pool-Seq, due to the method of filtering. In Pool-Seq, regions were filtered by MAPQ and read depth. However, in ddRAD-Seq, filtering included both the number of individuals contributing to a RADlocus, and the number of samples. Therefore, an increased chance of discarding regions which had decreased read alignment, due to allelic copies elsewhere in the genome, would have occurred.

Pool-Seq proved vastly more beneficial in these outbred, diverse field populations than ddRAD-Seq. By increasing the regional coverage of the genome, multiple narrow regions of genetic differentiation and potential IVM selection were identified. It also allowed increased sampling of the diversity within the population as a whole. Sample size was limited by the number of L3 available from the farms; however, increasing the number pooled would have led to decreased sampling per individual within the constraints of the sequencing budget. Thus, the actual number sampled would likely have remained very similar and would have reduced the benefit of including technical replicates. The coverage obtained following Pool-Seq was unfortunately less than expected, reducing the desired sample size. It was only after annotation of the genome using MTci2 RNA-Seq data that a clear explanation for this was identified as the draft genome itself. Although a considerable improvement upon the published *T. circumcincta* genome, the Tci2 draft genome remains in more than 8000 contigs. In addition, it was realised that it contains haplotypic copies of genes expected to be single copy. The full extent of this upon the present analysis is unknown, however very few genes within regions of genetic differentiation were identical when reads were aligned to both *T. circumcincta* genomes. Unfortunately,

similar issues exist in the published genome and the increased fragmentation (81,000 contigs) reduces the ability to determine regions of IVM selection.

To fully interpret the sequencing data generated in this study, it is vital that the Tci2 draft genome assembly is improved and that haplotypic contigs and allelic genes are reduced. The uncertainty relating to this issue is undoubtedly one of the greatest limitations of the study. It is also the point on which any future investigations rely for correctly identifying allele frequencies. Until this has been improved, or ascertained, there is less scope for improving or further developing GWAS. Using both PacBio long read sequences (Eid et al., 2009) and short read Illumina sequences, the current version of the Tci2 genome used in this PhD was assembled using FALCON-Unzip (Chin et al., 2016) and Redundans (Pryszcz and Gabaldón, 2016). The latter software aims to reduce haplotypic diversity within the genome (falsely duplicated contigs or genes) by percentage identity alignment of contigs one to another. In addition, the pre-library preparation technology, 10X genomics (Zheng et al., 2016), was employed to assist in identification of haplotypic regions for short read data. While this was an excellent start, the present draft assembly still retained artefactual duplication of many genes. As part of the BUG Consortium project, various methods are underway to improve the genome and overcome the high levels of haplotypic diversity inherent within *T. circumcincta* populations. Due to the small size of *T. circumcincta* adults and the low quantity of DNA that can be isolated from an individual worm, the DNA libraries sequenced for the current draft assembly of the Tci2 genome were produced from pools of up to 250,000 MTci2 L3. However, this has complicated the assembly process and may have introduced considerable diversity into the draft genome. The MTci2 isolate was originally obtained from the Central Veterinary Laboratories, Weybridge, in 2000 (Skuce et al., 2010), and no attempts have been made to inbreed, although the population may have experienced some changes over time following multiple passage within farm laboratory settings. In contrast, the original WSI *H. contortus* genome was produced following further inbreeding of the inbred susceptible isolate, MHco3(ISE), using a single adult of each sex to produce MHco3(ISE).N1, followed by sequencing of a single male worm (Laing et al., 2013; Sargison et al., 2017). This genome has since been improved upon using updated bioinformatics and third generation sequencing technologies (Tracey, *pers comm*). For this reason,

an attempt to inbreed the MTci2 population by the BUG Consortium, using offspring from single females as starter populations, is ongoing. In the future, to check whether a reduction in diversity has been achieved, microsatellites could be used to compare the inbred Tci2 with the MTci2 isolate population in a similar manner to previous work with *H. contortus* (Sargison et al., 2017).

There are various ways to identify multi-copy genes and some options could be applied to the currently available Tci2 genome. Firstly, the genome assembly was recently analysed (as part of this PhD) using BUSCO v3.0.2 (Benchmarking Universal Single-Copy Orthologues), to identify genes expected to be single copy, and determine current copy number (Simao et al., 2015; Waterhouse et al., 2017). This provided a better, genome-wide estimate of the current level of potential haplotypic duplication present within the Tci2 genome. Using the proteins annotated by AUGUSTUS, trained using the MTci2 RNA-Seq data, over 23% of genes in the BUSCO nematoda\_odb9 dataset were identified to be duplicated. The genome assembly was re-annotated using the *C. elegans* gene set within the nematoda\_odb9 dataset. Almost 20% of genes were identified to be duplicated, while 8% were not detected within the genome. Secondly, an attempt to reassemble the current genome could be made using the software Purge Haplotigs (Roach et al., 2018), which uses a slightly different method of identifying haplotypic sequences to Redundans (Pryszcz and Gabaldón, 2016). It first identifies potentially haplotypic contigs by the relative coverage of reads aligned to the genome. Following this, these contigs are compared to the rest of the genome, using Minimap2 (Li, 2018), and those appearing to be haplotypes of one another are identified as ‘haplotigs’. The shorter contig is removed, being retained as part of a separate ‘haplome’ and the remaining contigs are used to assemble the genome. Roach et al. (2018), found some improvement in genome assembly for three of the four species tested (a coral fungus, a grapevine and the Zebra finch), compared to assembly using Redundans. The fourth species, *Arabidopsis thaliana*, a model diploid plant, showed similar BUSCO statistics using either Purge Haplotigs or Redundans (Roach et al., 2018).

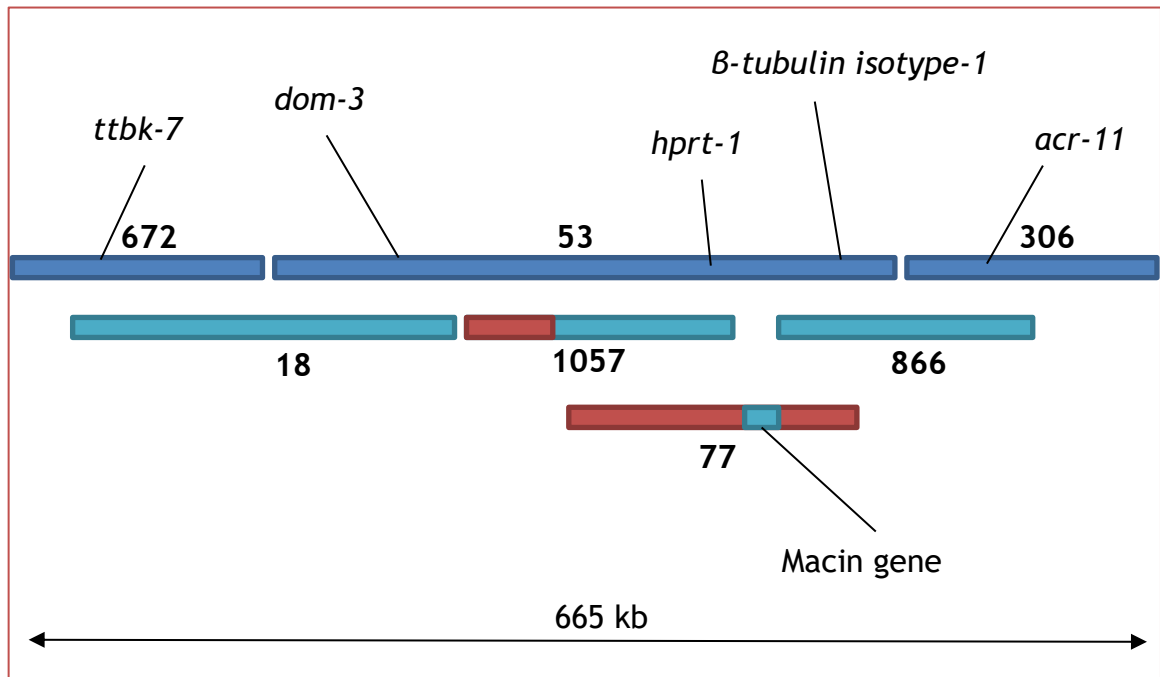
To improve the Tci2 genome assembly overall, the BUG Consortium is performing several additional experiments in collaboration with the WSI. The first is based on the successful sequencing and assembling of a 266 Mb mosquito genome,



using PacBio SMRT methodology, starting with amounts of DNA that were previously considered sub-optimal - 100 ng from a single individual (Kingan et al., 2019). The amount of DNA obtainable from a single worm is at best one tenth of this amount and it is unknown whether this will be successful, but if so, this method could provide excellent data. Secondly, the use of a technology known as Hi-C (Belton et al., 2012), which utilises the *in vivo* conformation of the genome to identify chromatin interactions, is also being trialled. In essence, the method allows identification of regions that are frequently spatially adjacent within the nucleus, indicating that they are likely to be close together within a linear genome (Belton et al., 2012). However, data obtained from such an experiment could also provide important functional information, as genes distantly located within a linear genome, but close within a cell, may be co-regulated. Thirdly, using a novel 10X genomics single cell method, an assembly using haploid genomes of *T. circumcincta* sperm cells may be possible. Unlike the original linked reads method (Zheng et al., 2016), which provided haplotype data, the newer method traps single cells, barcoding all DNA within the cell, before pooling DNA for sequencing (Andor et al., 2018). This is currently being trialled using *H. contortus* sperm, and, if successful, will be attempted with *T. circumcincta*. It should provide information as to the copy number of each gene within an individual male. Alternatively, individual unfertilised eggs could be used instead, with the caution that MHco3(ISE) unfertilised eggs have been identified to be diploid or aneuploid based on microsatellite analysis (Sargison, 2009). Triploid microsatellites were identified in Farm 1 *T. circumcincta* L3 individuals, and have been indicated previously in *T. circumcincta* populations (Grillo, 2005). They may reflect the presence of non-specific primer binding to a second region of the genome rather than triploidy of a locus. However, if triploids (whether euploidy, or more likely aneuploidy in nature) do exist, this may significantly complicate genome assembly and identification of gene copy number. Triploids may also provide an enhanced ability of a population to rapidly adapt to selection pressure so could play a role in the development of IVM resistance. For example, in human cancer cell lines, tetraploidy was associated with multi-drug resistance compared with parental cell lines (Kuznetsova et al., 2015). *H. contortus* polyploid individuals exist (Doyle et al., 2018), although it is unknown whether these are viable to infect sheep.

Whether a single locus, as suggested by *in vitro* testing of F1 and F2 generations of a *H. contortus* genetic cross (Le Jambre et al., 2000), or many loci, as suggested by work in *C. elegans* (Dent et al., 2000), are selected by IVM is not possible to determine using the current genome assemblies. Indeed, Doyle et al. (2019), demonstrated the marked improvement made to interpretation of NGS data by each refinement of the *H. contortus* genome assembly. Three studies have identified a large region on Chromosome V under IVM selection in *H. contortus* (Laing, *pers comm*, Doyle et al., 2019; Sallé et al., 2019). It is unclear whether the same large locus is syntenic in the *T. circumcincta* genome, however few homologues of the roughly 400 genes in the *H. contortus* locus appear to be under IVM selection in the *T. circumcincta* populations in this study. In the *H. contortus* genome, the large region of selection encompasses a region of reduced diversity and reduced recombination within the MHco3(ISE) population (Doyle et al., 2018; Doyle et al., 2019). In contrast, a similar large region of reduced haplotypic diversity, associated with IVM selection, was not detected within the data presented in the present study.

Nevertheless, when comparing *between* farms as part of this PhD, rather than within farms, a different set of genes appeared to be differentiated, including a single, large region surrounding the  $\beta$ -tubulin isotype-1 locus. This region correlated with regions identified by Choi et al. (2017) when the authors compared a triple resistant *T. circumcincta* isolate and an inbred susceptible isolate. A potentially homologous region on Chromosome I was also identified by Sallé et al. (2019), when geographically separated populations of *H. contortus* were compared. Importantly, Choi et al. (2017) could only identify a region of selection around the BZ locus as it extended on a single contig. By using both assemblies, as described in this thesis, the extension of this region to include genes on three PRJNA72569 contigs was possible (Figure 7.1). This demonstrated that the gene *acr-11*, identified by Choi et al. (2017) as potentially involved in LEV resistance, lies very close to  $\beta$ -tubulin isotype-1. As such, *acr-11* may be selected by genetic hitch-hiking with the  $\beta$ -tubulin isotype-1 gene (related to selection by BZ), rather than associated with LEV sensitivity itself. Many of the genes identified in this study are likely indirectly selected by IVM treatment. Although closely linked genes could be used as genetic markers of IVM resistance, recombination could render them ineffective.



**Figure 7.1:** Relative position of contigs within the *Teladorsagia circumcincta* genome. The schematic shows the relationship between three PRJNA72569 contigs (672, 53 and 306, dark blue) and four Tci2 contigs (18, 1057, 77 and 866). These positions were determined using evidence of differentiation (light blue) over the Tci2 regions when comparing between Farm 2 and Farm 3 populations, and the identification of their equivalent genes in the PRJNA72569 genome. Note that contigs 77 and 1057 also contain undifferentiated regions (red). These regions contained genes which did not map to PRJNA72569 contig 53. A selection of genes are annotated.

Why might a wide region of reduced diversity arise around the BZ locus, but does not appear to have occurred around the (unknown) IVM locus? Several reasons are possible. Recombination rate varies both between and along chromosomes (Doyle et al., 2018), and between species (Stapley et al., 2017). The recombination rate surrounding the BZ locus in the *T. circumcincta* genome may be reduced compared with the IVM locus. Thus, the potential for linked loci to reach fixation is higher (Smith and Haigh, 1974). The difference in dominance of the resistance associated SNPs may have increased selection of BZ homozygous resistant individuals (an incompletely recessive trait in *H. contortus* (Sangster et al., 1998)), compared with IVM, (a dominant trait in *T. circumcincta* (Sutherland et al., 2002) and *H. contortus* (Le Jambre et al., 2000)). The frequency of selection may also vary: farmers used to be told to treat neo-suppressively, which can rapidly select for just a few haplotypes, as was observed when *Anopheles funestus* was subjected to intense selection by pyrethroids (Barnes et al., 2017), but fewer will do so now. BZ is an older drug and is more commonly used in this way even today for *N. battus*.

Missing from current knowledge is how anthelmintics from different classes may influence selection upon other anthelmintic resistance mechanisms. For example, one study noted apparent selection upon BZ resistance associated SNPs by IVM when IVM resistant isolates were compared with IVM susceptible isolates of *H. contortus*, including comparison of derived strains with their parental isolate (de Lourdes Mottier and Prichard, 2008). Although one IVM resistant strain with BZ associated SNPs identified was derived from a BZ naïve isolate, with no individuals containing F200Y SNPs identified in this parental isolate, only 27 worms were genotyped from each. Indeed, individuals within the parental isolate population were noted to contain F167Y SNPs. It is possible that P200 BZ SNPs were already present within the parental isolate population and that the apparent selection observed related to the presence of dual-resistant individuals. To distinguish between the selection of dual-resistant *individuals*, and selection by one anthelmintic upon the specific genetic locus conferring *in vivo* resistance to a second drug class, is important. The former is affected by management and is largely dependent on past treatment history. The latter might enhance selection of resistance to both anthelmintic classes, independent of management factors, each time the anthelmintic is used. This selection could occur either by the same mechanism for both drugs, or via genetic linkage (hitch-hiking) between two separate mechanisms (Smith and Haigh, 1974). Until we know the genetic basis of IVM resistance and observe if this is linked with other anthelmintic resistance loci, we cannot fully answer these questions. Despite this, the present study has not found evidence of selection upon BZ associated SNPs by IVM. Indeed, differentiation over the  $\beta$ -tubulin isotype-1 locus was only identified when different farm populations were compared, which may relate to decreased within-population nucleotide diversity at this locus. Similarly, P-glycoproteins have been suggested to be an IVM resistance mechanism, but apart from a tentative link with *pgp-3*, this study failed to identify an association between P-gps and IVM resistance. It is possible that this gene family presents a multi-drug resistance mechanism, rather than an IVM specific mechanism. For example, P-gps also appear to be associated with exposure to MPTL (Raza et al., 2016a), resistance to LEV (Raza et al., 2016b), and resistance to BZ (Blackhall et al., 2008) in *H. contortus* isolates. As mentioned previously, the sample populations could have contained dual- or

multi-drug resistant individuals so it is also possible that any P-gp mediated resistance was already present in the pre-treatment populations due to an involvement in, for example, BZ resistance. Conversely, some genes identified in this study (e.g. AChRs) could be associated with other anthelmintics used on these farms.

Future adaptations to genome sequencing methodologies, as described above, will allow improvements to the genome, and, at that point, reads generated in this study will be re-aligned and re-analysed.  $F_{ST}$  windows could be filtered by SNP density, excluding those with very few or a large number of SNPs as was used by Cheeseman et al. (2015) to reduce false positives within their Pool-Seq data. This could help to reduce noise within the Farm 2 data in particular. Rather than just using smoothed  $F_{ST}$ , the significance of any allele frequency difference could be smoothed also (Kofler et al., 2011b), and this additional statistic used to investigate results. To allow the use of a larger window size, the new genome could be subdivided by contig size, dependent on the state of the assembly, into two smaller 'genomes' following read alignment. In this way longer contigs could be separately examined using larger sliding windows, while shorter contigs could be analysed with 5 or 10 kb windows. Coverage along the genome in the present study was found to vary, even within a contig. To assist in a reasonable sub-sampling protocol, the average coverage of genes identified as single copy by BUSCO (Simao et al., 2015) could be determined, particularly over exons. Regions of coverage similar to single gene coverage could be analysed separately to those of lower or higher coverage, which may help with allele frequency sampling bias.

Even with improvement of the genome, it is likely that there will still be a considerable number of genes identified as potential markers of IVM resistance. For example, the region in *H. contortus* identified on Chromosome V contains a considerable number of genes (Doyle et al., 2019; Sallé et al., 2019), despite expectations of a single dominant locus conferring resistance (Le Jambre et al., 2000). Even in relatively outbred clinical isolates of *O. volvulus* (Doyle et al., 2017), multiple genes were identified as potentially under selection, although in this species resistance may be polygenic. It is also not necessarily the case that the resistance mechanism will be identified within the highest  $F_{ST}$  window, or

with a single, strongly selected SNP. For example, Cheeseman et al. (2015) identified the *kelch* gene, expected to be a key determinant of artemisinin resistance in *P. falciparum* (Ariey et al., 2014) within their main locus of selection by  $F_{ST}$ . Yet other nearby genes had higher  $F_{ST}$  values than the *kelch* gene, because multiple haplotypes of the *kelch* gene, containing different non-synonymous SNPs, were associated with resistance phenotypes (Cheeseman et al., 2015). Each of these SNPs contributed only a small  $F_{ST}$  value. Similarly, although BZ resistance is primarily conferred by changes in the *ben-1* gene in *C. elegans*, multiple variants conferring a resistance phenotype are possible (Hahnel et al., 2018). If multiple SNPs are capable of conferring IVM resistance in *T. circumcincta*, and are present within these farm populations, then each SNP may have only a small  $F_{ST}$  value relative to more common SNPs nearby. Therefore, although we have identified many interesting genes within regions of genetic differentiation by  $F_{ST}$ , and these have a rank based on their  $F_{ST}$  value, the rank may not correlate with their importance in IVM resistance. Some, such as those predicted to be LGIC homologues, have lower ranked values, but are of greater interest, than other genes predicted to be related to the cuticle, for example.

Targeted sequencing of additional populations may eliminate genes present within this study due to genetic hitch-hiking and help to identify markers common to many UK farms, useful for diagnostic testing or research. Nevertheless, multiple markers may exist, with some important on one farm, and different markers useful to a second flock as is noted with different SNPs within the  $\beta$ -tubulin isotype-1 gene and BZ resistance (Redman et al., 2015), and with different haplotypes of *Tci-mptl-1* and MPTL resistance (Turnbull et al., 2019). Unfortunately, sequencing of individual worms is limited by the total amount of DNA available (< 1 ng from a *T. circumcincta* L3). Array technology (Gabriel et al., 2009; Bumgarner, 2013) could allow the development of a customised SNP panel, as has been produced for other parasites, e.g. *P. falciparum* (Tan et al., 2011). The panel would include selected and neutral SNPs and could be used to test many farm populations at relatively low cost. While the quantity of DNA required for a SNP array is much greater than that obtained from a single worm, population-level data are likely to be more useful for diagnostic purposes. Therefore, one option would be to use Illumina

GoldenGate technology, which allows pooling of barcoded individuals for SNP arrays and has been used to study insecticide resistance (Weetman et al., 2018). Other studies have used whole genome amplification of parasites to obtain sufficient DNA for array based technologies (Doyle et al., 2017), however this could be prohibitively expensive. The use of targeted sequencing of specific genomic regions of interest using MiSeq with pools of individuals is showing great promise within the ruminant nematode community. It has been used both to identify species abundance within a sample (Avramenko et al., 2015) and to detect BZ resistance associated SNPs (Avramenko et al., 2019). With this technology it is possible to barcode and pool up to 96 samples for a single sequencing reaction, with a sequencing cost in the region of £2000. Based on gDNA quantities obtained for Pool-Seq in the present study, it should be possible to obtain sufficient quantities of DNA from pools of 200 L3 to perform 35 MiSeq reactions.

If IVM resistance carries a fitness cost then individuals may be penalised in the absence of drug as was noted in *Aedes aegypti* following removal of pyrethroids (Chang et al., 2012; Vera-Maloof et al., 2019). A study by Leathwick et al. (2015) suggested some reduction in IVM resistance of *T. circumcincta* field populations after five years without IVM treatment. However, improvement year-on-year for each farm was inconsistent, and to date, no studies have confirmed a fitness cost of anthelmintic resistance in ruminant strongyles (Hodgkinson et al., 2019). Nevertheless, for a long time, the use of anthelmintics by farmers has been frequent (Burgess et al., 2012; McMahon et al., 2013a), and it is possible that fitness costs have not been observed due to the relative advantage conferred to parasites by this management practice. Indeed, if dual-resistance is common, as might be expected with frequent dosing using alternating or combination products, then a complete absence of anthelmintics might be necessary to observe fitness costs.

The ready availability of highly efficacious anthelmintics has allowed farmers to maintain production with sheep that otherwise would be susceptible to parasitic gastroenteritis. However, more recently, with increasing anthelmintic resistance, UK farmers are actively breeding for resistant or resilient animals (Fairlie-Clarke et al., 2019). In addition, climate is changing, affecting strongyle

communities (Kenyon et al., 2009b; Rose et al., 2016). In the future, these management and climatic changes may identify hitherto unrealised fitness costs of anthelmintic resistance in parasite populations. To exploit these, or actively maintain a susceptible population *in refugia*, we need to know the genetic basis of IVM resistance to identify optimal management practices. Once IVM resistance markers are available, it may be possible to use targeted re-sequencing of populations, for example using MiSeq, to monitor changes in IVM resistance markers in a population over time. Experiments could investigate the impact of *refugia*, of management practices such as targeted selective treatment (Kenyon et al., 2009a; Kenyon et al., 2013), and the effect of other anthelmintic classes upon selection of IVM resistance markers (or *vice versa*). The latter may be of particular importance for MOX treatments, which may or may not select for IVM resistance (Sargison et al., 2012). Using MiSeq, multiple loci could be amplified, using the same *T. circumcincta* samples. Although moderately expensive, and with an intensive initial labour input, such experiments would produce a huge amount of novel and interesting data and would have the potential to answer several research questions at once.

## 7.1 Conclusion

Identifying markers of IVM resistance proved extremely challenging. It is possible that, due to the draft state of the Tci2 genome, regions of IVM selection may have been missed, while other regions have been falsely implicated.

Nevertheless, an interesting and compelling set of neuronal genes were identified within genetically differentiated loci, as were others involved in metabolism generally and lipid metabolism specifically. Other gene types were also common in the analysis, including genes with suggested regulatory functions and transposable elements. A recent paper implicated changes in the function of a nematode nuclear hormone receptor, *Cel-nhr-8*, in IVM tolerance in *C. elegans* (Ménez et al., 2019), and it may be that in the field, certain genes are selected for tolerance rather than resistance. IVM causes pharyngeal paralysis and thus induces a period of starvation in larvae (Gill et al., 1995). If the same phenomenon occurs in partially resistant adult worms, but worms are not directly killed by IVM, genes related to metabolism and fat storage may be crucial in worm survival (Laing et al., 2012). A number of genes appeared to be



associated with pharyngeal structures and chemosensation. It may be that IVM indirectly induces inhibition of feeding via chemosensory pathways, or that resistance mechanisms involve reduced uptake of IVM via the oral route. Experimental studies have previously suggested IVM resistance to be dominant in *T. circumcincta* (Sutherland et al., 2002) and the findings of this PhD support that. No single region appeared strongly differentiated by  $F_{ST}$  in comparison to BZ resistance markers identified in other studies (Laing, *pers comm*, Choi et al., 2017; Sallé et al., 2019), or the recessively inherited oxamniquine resistance marker (Valentim et al., 2013). Although the analysis has proven challenging, and future re-analysis of the data using improved reference genomes will be beneficial, some excellent field population NGS resources have been produced and protocols suitable for use in future studies have been validated.

## Appendices

All appendices can be found on the USB stick within the envelope at the end of this thesis.

### **Appendix 1 Arlequin parameters chosen.docx**

The parameters selected for the microsatellite analysis.

### **Appendix 2 Bioinformatics Scripts.docx**

The relevant bioinformatics scripts used. These will be uploaded onto GitHub in the future.

### **Appendix 3 ddRAD-Seq de-multiplexed summary statistics.docx**

The Stacks *process\_radtags* summary statistics for the Farm 2 individuals sequenced using ddRAD-Seq.

### **Appendix 4 ddRAD-Seq data (Farm 2).xlsx**

The  $F_{ST}$ ,  $F'_{ST}$  and  $D_{EST}$  results from the ddRAD-Seq analysis. Only those results greater than each threshold are included.

### **Appendix 5 Pool-Seq High 10 kb windows identified on both Farms 2 and 3.xlsx**

The 10 kb regions identified as genetically differentiated on both farms in the Pool-Seq analysis.

### **Appendix 6 Pool-Seq High 10 kb windows identified in the Farm 2 analysis.xlsx**

### **Appendix 7 Pool-Seq High 10 kb windows identified in the Farm 3 analysis.xlsx**

These contain the Pool-Seq data ( $F_{ST}$ , genes and predicted proteins) for each farm analysis. Only those 10 kb windows greater than the thresholds are included.

**Appendix 8 Comparison of ddRAD-Seq Farm 2 SNPs with Pool-Seq Farm 2.xlsx**

**Appendix 9 Comparison of ddRAD-Seq Farm 2 SNPs with Pool-Seq Farm 3.xlsx**

These contain the results after comparing the ddRAD-Seq Farm 2 data with the Pool-Seq data.

**Appendix 10 Farm 2 GO analysis BIOMART output.xlsx**

**Appendix 11 Farm 3 GO analysis BIOMART output.xlsx**

The information extracted using BIOMART is presented, with the four gene categories shown (neuronal, lipid, male and female related terms).

**Appendix 12 Comparison of Pool-Seq with Choi et al 2017.xlsx**

Includes the results of the analysis comparing the Pool-Seq data from each farm with the results presented by Choi et al., 2017.

**Appendix 13 Comparison of Pool-Seq with *H. contortus* data provided by Dr R Laing.xlsx**

Includes the results of the analysis comparing the Pool-Seq data from each farm with the *H. contortus* MHco3(ISE)xMHco18(UGA) results provided by Dr R Laing (unpublished data).

**Appendix 14 Comparison of Pool-Seq with *O. volvulus* Doyle et al 2017.xlsx**

Includes the results of the analysis comparing the Pool-Seq data from each farm with the results presented by Doyle et al., 2017.

**Appendix 15 CNV logfiles for pgp.docx**

Includes additional information regarding the identification of the *pgp-3* and *pgp-9* Tci2 genes.

**Appendix 16 Published paper (open access)**

McIntyre, J., et al. (2018). "Hidden in plain sight - Multiple resistant species within a strongyle community." *Vet Parasitol* 258: 79-87.

Note, this is not included in the electronic appendices on Enlighten, but can be found using the following DOI: <https://doi.org/10.1016/j.vetpar.2018.06.012>

## List of References

- Abbott, K.A., Taylor, M.A., Stubbings, L.A., 2012. Sustainable Worm Control Strategies for Sheep: A Technical Manual for Veterinary Surgeons and Advisers. SCOPS 4th edition.
- Abongwa, M., Buxton, S.K., Robertson, A.P., Martin, R.J., 2016. Curiouser and Curiouser: The Macrocyclic Lactone, Abamectin, Is also a Potent Inhibitor of Pyrantel/Tribendimidine Nicotinic Acetylcholine Receptors of Gastro-Intestinal Worms. PLoS One 11, e0146854.
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46, W537-W544.
- AHDB 2017. Exploring the implications of BREXIT for the agriculture and horticulture in Scotland.
- AHDB 2018. UK Sheep Yearbook 2018 (Agricultural and Horticultural Development Board (AHDB) Lamb).
- Ahluwalia, J.S., Charleston, W.A., 1974. Studies on the development of the free-living stages of *Cooperia curticei*. N Z Vet J 22, 191-195.
- Ahluwalia, J.S., Charleston, W.A., 1975. Studies on the pathogenicity of *Cooperia curticei* for sheep. N Z Vet J 23, 197-199.
- Ahmed, A.M., Sebastiano, S.R., Sweeney, T., Hanrahan, J.P., Glynn, A., Keane, O.M., Mukhopadhyaya, A., Thornton, K., Good, B., 2015. Breed differences in humoral and cellular responses of lambs to experimental infection with the gastrointestinal nematode *Teladorsagia circumcincta*. Vet Res 46, 8-8.
- Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12, R18.
- Alary, V., Corniaux, C., Gautier, D., 2011. Livestock's Contribution to Poverty Alleviation: How to Measure It? World Dev 39, 1638-1648.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J Mol Biol 215, 403-410.
- Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., Lander, E.S., 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nat 407, 513-516.
- Alvarez-Sanchez, M.A., Perez-Garcia, J., Cruz-Rojo, M.A., Rojo-Vazquez, F.A., 2005a. Real time PCR for the diagnosis of benzimidazole resistance in trichostrongylids of sheep. Vet Parasitol 129.
- Alvarez-Sanchez, M.A., Perez Garcia, J., Bartley, D., Jackson, F., Rojo-Vazquez, F.A., 2005b. The larval feeding inhibition assay for the diagnosis of nematode anthelmintic resistance. Exp Parasitol 110, 56-61.
- Alvarez, L., Suarez, G., Ceballos, L., Moreno, L., Canton, C., Lifschitz, A., Mate, L., Ballent, M., Virkel, G., Lanusse, C., 2015. Integrated assessment of ivermectin pharmacokinetics, efficacy against resistant *Haemonchus contortus* and P-glycoprotein expression in lambs treated at three different dosage levels. Vet Parasitol 210, 53-63.
- Aminetzach, Y.T., Macpherson, J.M., Petrov, D.A., 2005. Pesticide Resistance via Transposition-Mediated Adaptive Gene Truncation in *Drosophila*. Science 309, 764-767.

- Anderson, E.C., Skaug, H.J., Barshis, D.J., 2014. Next-generation sequencing for molecular ecology: a caveat regarding pooled samples. *Mol Ecol* 23, 502-512.
- Andor, N., Lau, B.T., Catalanotti, C., Kumar, V., Sathe, A., Belhocine, K., Wheeler, T.D., Price, A.D., Song, M., Stafford, D., Bent, Z., DeMare, L., Hepler, L., Jett, S., Lin, B.K., Maheshwari, S., Makarewicz, A.J., Rahimi, M., Sawhney, S.S., Sauzade, M., Shuga, J., Sullivan-Bibee, K., Weinstein, A., Yang, W., Yin, Y., Kubit, M.A., Chen, J., Grimes, S.M., Suarez, C.J., Poultides, G.A., Schnall-Levin, M., Bharadwaj, R., Ji, H.P., 2018. Joint single cell DNA-Seq and RNA-Seq of gastric cancer reveals subclonal signatures of genomic instability and gene expression. *bioRxiv*, 445932.
- Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet* 17, 81-92.
- APHA 2018. Veterinary Investigation Diagnosis Analysis (VIDA) Annual report 2017.
- Arena, J.P., Liu, K.K., Paress, P.S., Frazier, E.G., Cully, D.F., Mrozik, H., Schaeffer, J.M., 1995. The mechanism of action of avermectins in *Caenorhabditis elegans*: correlation between activation of glutamate-sensitive chloride current, membrane binding, and biological activity. *J Parasitol* 81, 286-294.
- Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C.M., Bout, D.M., Menard, S., Rogers, W.O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.C., Fairhurst, R.M., Benoit-Vical, F., Mercereau-Puijalon, O., Menard, D., 2014. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nat* 505, 50-55.
- Arnold, B., Corbett-Detig, R.B., Hartl, D., Bomblies, K., 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol* 22, 3179-3190.
- Avramenko, R.W., Redman, E.M., Lewis, R., Yazwinski, T.A., Wasmuth, J.D., Gilleard, J.S., 2015. Exploring the Gastrointestinal "Nemabiome": Deep Amplicon Sequencing to Quantify the Species Composition of Parasitic Nematode Communities. *PLoS One* 10, e0143559.
- Avramenko, R.W., Redman, E.M., Melville, L., Bartley, Y., Wit, J., Queiroz, C., Bartley, D.J., Gilleard, J.S., 2019. Deep amplicon sequencing as a powerful new tool to screen for sequence polymorphisms associated with anthelmintic resistance in parasitic nematode populations. *Int J Parasitol* 49, 13-26.
- Awadzi, K., Attah, S.K., Addy, E.T., Opoku, N.O., Quartey, B.T., Lazdins-Helds, J.K., Ahmed, K., Boatin, B.A., Boakye, D.A., Edwards, G., 2004. Thirty-month follow-up of sub-optimal responders to multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Ann Trop Med Parasitol* 98, 359-370.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3, e3376.
- Baker, N.F., Cook, E.F., Douglas, J.R., Cornelius, C.E., 1959. The pathogenesis of trichostrongyloid parasites. III. Some physiological observations in lambs suffering from acute parasitic gastroenteritis. *J Parasitol* 45.

- Balic, A., Bowles, V.M., Liu, Y.S., Meeusen, E.N.T., 2003. Local immune responses in sensitized sheep following challenge infection with *Teladorsagia circumcincta*. *Parasite Immunol* 25, 375-381.
- Ballesteros, C., Tritten, L., O'Neill, M., Burkman, E., Zaky, W.I., Xia, J., Moorhead, A., Williams, S.A., Geary, T.G., 2016. The Effects of Ivermectin on *Brugia malayi* Females *In Vitro*: A Transcriptomic Approach. *PLoS Negl Trop Dis* 10, e0004929.
- Barbu, E., Lee, K.Y., Wahl, R., 1956. Content of purine and pyrimidine base in desoxyribonucleic acid of bacteria. *Ann Inst Pasteur (Paris)* 91, 212-224.
- Barnes, K.G., Weedall, G.D., Ndula, M., Irving, H., Mzihalowa, T., Hemingway, J., Wondji, C.S., 2017. Genomic Footprints of Selective Sweeps from Metabolic Resistance to Pyrethroids in African Malaria Vectors Are Driven by Scale up of Insecticide-Based Vector Control. *PLoS Genet* 13, e1006539.
- Barnes, T.M., Kohara, Y., Coulson, A., Hekimi, S., 1995. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genet* 141, 159-179.
- Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stromberg, M.P., Marth, G.T., 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27, 1691-1692.
- Barrère, V., Alvarez, L., Suarez, G., Ceballos, L., Moreno, L., Lanusse, C., Prichard, R.K., 2012. Relationship between increased albendazole systemic exposure and changes in single nucleotide polymorphisms on the  $\beta$ -tubulin isotype 1 encoding gene in *Haemonchus contortus*. *Vet Parasitol* 186, 344-349.
- Barrère, V., Beech, R.N., Charvet, C.L., Prichard, R.K., 2014. Novel assay for the detection and monitoring of levamisole resistance in *Haemonchus contortus*. *Int J Parasitol* 44, 235-241.
- Barrett, M., Jackson, F., Huntley, J.F., 1998. Pathogenicity and immunogenicity of different isolates of *Teladorsagia circumcincta*. *Vet Parasitol* 76, 95-104.
- Bartley, D.J., Jackson, E., Johnston, K., Coop, R.L., Mitchell, G.B., Sales, J., Jackson, F., 2003. A survey of anthelmintic resistant nematode parasites in Scottish sheep flocks. *Vet Parasitol* 117, 61-71.
- Bartley, D.J., Jackson, F., Jackson, E., Sargison, N., 2004. Characterisation of two triple resistant field isolates of *Teladorsagia* from Scottish lowland sheep farms. *Vet Parasitol* 123, 189-199.
- Bartley, D.J., Donnan, A.A., Jackson, E., Sargison, N., Mitchell, G.B.B., Jackson, F., 2006. A small scale survey of ivermectin resistance in sheep nematodes using the faecal egg count reduction test on samples collected from Scottish sheep. *Vet Parasitol* 137.
- Bartley, D.J., McAllister, H., Bartley, Y., Dupuy, J., Menez, C., Alvinerie, M., Jackson, F., Lespine, A., 2009. P-glycoprotein interfering agents potentiate ivermectin susceptibility in ivermectin sensitive and resistant isolates of *Teladorsagia circumcincta* and *Haemonchus contortus*. *Parasitol* 136, 1081-1088.
- Bartley, D.J., Devin, L., Nath, M., Morrison, A.A., 2015. Selection and characterisation of monepantel resistance in *Teladorsagia circumcincta* isolates. *Int J Parasitol Drugs Drug Resist* 5, 69-76.
- Bartley, D.J., Meslé, M., Donegan, H., Devin, L., Morrison, A.A., 2016. Phenotypic assessment of the ovicidal activity of monepantel and monepantel sulfone on gastro-intestinal nematode eggs. *Vet Parasitol* 220, 87-92.

- Bartley, D.J., Hamer, K., Andrews, L., Sargison, N.D., Morrison, A.A., 2019. Multigeneric resistance to monepantel on a UK sheep farm. *Vet Parasitol*: X, 100003.
- Beech, R., Levitt, N., Cambos, M., Zhou, S., Forrester, S.G., 2010. Association of ion-channel genotype and macrocyclic lactone sensitivity traits in *Haemonchus contortus*. *Mol Biochem Parasitol* 171, 74-80.
- Beech, R.N., Prichard, R.K., Scott, M.E., 1994. Genetic variability of the beta-tubulin genes in benzimidazole-susceptible and -resistant strains of *Haemonchus contortus*. *Genet* 138, 103.
- Beech, R.N., Skuce, P., Bartley, D.J., Martin, R.J., Prichard, R.K., Gilleard, J.S., 2011. Anthelmintic resistance: markers for resistance, or susceptibility? *Parasitol* 138, 160-174.
- Beesley, N.J., Williams, D.J., Paterson, S., Hodgkinson, J., 2017. *Fasciola hepatica* demonstrates high levels of genetic diversity, a lack of population structure and high gene flow: possible implications for drug resistance. *Int J Parasitol* 47, 11-20.
- Beissinger, T.M., Rosa, G.J.M., Kaeppler, S.M., Gianola, D., de Leon, N., 2015. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genet Sel Evol* 47, 30-30.
- Bekelaar, K., Waghorn, T., Tavendale, M., McKenzie, C., Leathwick, D., 2018. Carbon dioxide is an absolute requirement for exsheathment of some, but not all, abomasal nematode species. *Parasitol Res* 117, 3675-3678.
- Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods (San Diego, Calif)* 58, 268-276.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara E Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.-D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C.,



- Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczky, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nat* 456, 53-59.
- Bergamaschi, D., Samuels, Y., O'Neil, N.J., Trigiante, G., Crook, T., Hsieh, J.-K., O'Connor, D.J., Zhong, S., Campargue, I., Tomlinson, M.L., Kuwabara, P.E., Lu, X., 2003. iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. *Nat Genet* 33, 162-167.
- Beveridge, I., Pullman, A.L., Phillips, P.H., Martin, R.R., Barelds, A., Grimson, R., 1989. Comparison of the effects of infection with *Trichostrongylus colubriformis*, *T. vitrinus* and *T. rugatus* in merino lambs. *Vet Parasitol* 32, 229-245.
- Bird, C.E., Karl, S.A., Smouse, P.E., Toonen, R.J., 2011. Detecting and measuring genetic differentiation, In: Held, C., Koenemann, S., Schubart, C.D. (Eds.) *Phylogeography and Population Genetics in Crustacea*. Boca Raton, FL: CRC press, pp. 31-55.
- Bishop, S.C., Stear, M.J., 2000. The use of a gamma-type function to assess the relationship between the number of adult *Teladorsagia circumcincta* and total egg output. *Parasitol* 121 (Pt 4), 435-440.
- Bisset, S.A., 2007. The Genetic Basis of Multiple-anthelmintic Resistance in *Teladorsagia circumcincta*, a Gastrointestinal nematode Parasite of Sheep and Goats.
- Bisset, S.A., Knight, J.S., Bouchet, C.L., 2014. A multiplex PCR-based method to identify strongylid parasite larvae recovered from ovine faecal cultures and/or pasture samples. *Vet Parasitol* 200, 117-127.
- Blackhall, W.J., Pouliot, J.F., Prichard, R.K., Beech, R.N., 1998. *Haemonchus contortus*: selection at a glutamate-gated chloride channel gene in ivermectin- and moxidectin-selected strains. *Exp Parasitol* 90, 42-48.
- Blackhall, W.J., Prichard, R.K., Beech, R.N., 2003. Selection at a gamma-aminobutyric acid receptor gene in *Haemonchus contortus* resistant to avermectins/milbemycins. *Mol Biochem Parasitol* 131, 137-145.
- Blackhall, W.J., Prichard, R.K., Beech, R.N., 2008. P-glycoprotein selection in strains of *Haemonchus contortus* resistant to benzimidazoles. *Vet Parasitol* 152, 101-107.
- Blanchard, A., Guegnard, F., Charvet, C.L., Crisford, A., Courtot, E., Sauve, C., Harmache, A., Duguet, T., O'Connor, V., Castagnone-Sereno, P., Reaves, B., Wolstenholme, A.J., Beech, R.N., Holden-Dye, L., Neveu, C., 2018. Deciphering the molecular determinants of cholinergic anthelmintic sensitivity in nematodes: When novel functional validation approaches highlight major differences between the model *Caenorhabditis elegans* and parasitic species. *PLoS Pathog* 14, e1006996.
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., Team, t.G., 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26, 1783-1785.

- Boag, B., Thomas, R.J., 1977. Epidemiological studies on gastro-intestinal nematode parasites of sheep: the seasonal number of generations and succession of species. *Res Vet Sci* 22, 62-67.
- Boitard, S., Schlotterer, C., Nolte, V., Pandey, R.V., Futschik, A., 2012. Detecting selective sweeps from pooled next-generation sequencing samples. *Mol Biol Evol* 29, 2177-2186.
- Boulin, T., Gielen, M., Richmond, J.E., Williams, D.C., Paoletti, P., Bessereau, J.-L., 2008. Eight genes are required for functional reconstitution of the *Caenorhabditis elegans* levamisole-sensitive acetylcholine receptor. *P Natl A Sci* 105, 18590-18595.
- Boulin, T., Fauvin, A., Charvet, C.L., Cortet, J., Cabaret, J., Bessereau, J.L., Neveu, C., 2011. Functional reconstitution of *Haemonchus contortus* acetylcholine receptors in *Xenopus oocytes* provides mechanistic insights into levamisole resistance. *Br J Pharmacol* 164, 1421-1432.
- Bourguinat, C., Lee, A.C., Lizundia, R., Blagburn, B.L., Liotta, J.L., Kraus, M.S., Keller, K., Epe, C., Letourneau, L., Kleinman, C.L., Paterson, T., Gomez, E.C., Montoya-Alonso, J.A., Smith, H., Bhan, A., Peregrine, A.S., Carmichael, J., Drake, J., Schenker, R., Kaminsky, R., Bowman, D.D., Geary, T.G., Prichard, R.K., 2015. Macrocyclic lactone resistance in *Dirofilaria immitis*: Failure of heartworm preventives and investigation of genetic markers for resistance. *Vet Parasitol* 210, 167-178.
- Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., Mager, D.L., Feschotte, C., 2018. Ten things you should know about transposable elements. *Genome Biol* 19, 199.
- Britt, D.P., 1982. Benzimidazole-resistant nematodes in Britain. *Vet Rec* 110, 343-344.
- Bröer, S., Gether, U., 2012. The solute carrier 6 family of transporters. *Br J Pharmacol* 167, 256-278.
- Brunsdon, R.V., 1966. A comparison of the spring-rise phenomenon in the faecal nematode-egg counts of housed sheep with that of sheep grazing infective pasture. *N Z Vet J* 14, 145-151.
- Bumgarner, R., 2013. Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol* Chapter 22, Unit-22.21.
- Burgess, C.G., Bartley, Y., Redman, E., Skuce, P.J., Nath, M., Whitelaw, F., Tait, A., Gilleard, J.S., Jackson, F., 2012. A survey of the trichostrongylid nematode species present on UK sheep farms and associated anthelmintic control practices. *Vet Parasitol* 189, 299-307.
- Cabaret, J., Gasnier, N., Jacquiet, P., 1998. Faecal egg counts are representative of digestive-tract strongyle worm burdens in sheep and goats. *Parasite* 5, 137-142.
- Calvete, C., Ferrer, L.M., Lacasta, D., Calavia, R., Ramos, J.J., Ruiz-de-Arkaute, M., Uriarte, J., 2014. Variability of the egg hatch assay to survey benzimidazole resistance in nematodes of small ruminants under field conditions. *Vet Parasitol* 203, 102-113.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
- Campbell, W.C., Fisher, M.H., Stapley, E.O., Albers-Schonberg, G., Jacob, T.A., 1983. Ivermectin: a potent new antiparasitic agent. *Science* 221, 823-828.
- Campbell, W.C., 1985. Ivermectin: an update. *Parasitol Today* 1, 10-16.

- Campbell, W.C., 1991. Ivermectin as an antiparasitic agent for use in humans. *Annu Rev Microbiol* 45, 445-474.
- Cariou, M., Duret, L., Charlat, S., 2016. How and how much does RAD-seq bias genetic diversity estimates? *BMC Evol Biol* 16, 240.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol* 22, 3124-3140.
- Catchen, J.M., Hohenlohe, P.A., Bernatchez, L., Funk, W.C., Andrews, K.R., Allendorf, F.W., 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Mol Ecol Resour* 17, 362-365.
- Cawthorne, R.J., Cheong, F.H., 1984. Prevalence of anthelmintic resistant nematodes in sheep in south-east England. *Vet Rec* 114, 562-564.
- Chabala, J.C., Mrozik, H., Tolman, R.L., Eskola, P., Lusi, A., Peterson, L.H., Woods, M.F., Fisher, M.H., Campbell, W.C., Egerton, J.R., Ostlind, D.A., 1980. Ivermectin, a new broad-spectrum antiparasitic agent. *J Med Chem* 23, 1134-1136.
- Chalasani, S.H., Chronis, N., Tsunozaki, M., Gray, J.M., Ramot, D., Goodman, M.B., Bargmann, C.I., 2007. Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nat* 450, 63.
- Chalfie, M., Sulston, J., 1981. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Dev Biol* 82, 358-370.
- Chalfie, M., Wolinsky, E., 1990. The identification and suppression of inherited neurodegeneration in *Caenorhabditis elegans*. *Nat* 345, 410-416.
- Chang, C., Huang, X.-Y., Chang, P.-C., Wu, H.-H., Dai, S.-M., 2012. Inheritance and stability of sodium channel mutations associated with permethrin knockdown resistance in *Aedes aegypti*. *Pestic Biochem Phys* 104, 136-142.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4.
- Chaudhry, U., Redman, E.M., Raman, M., Gilleard, J.S., 2015. Genetic evidence for the spread of a benzimidazole resistance mutation across southern India from a single origin in the parasitic nematode *Haemonchus contortus*. *Int J Parasitol* 45, 721-728.
- Cheeseman, I.H., Miller, B.A., Nair, S., Nkhoma, S., Tan, A., Tan, J.C., Al Saai, S., Phyo, A.P., Moo, C.L., Lwin, K.M., McGready, R., Ashley, E., Imwong, M., Stepniewska, K., Yi, P., Dondorp, A.M., Mayxay, M., Newton, P.N., White, N.J., Nosten, F., Ferdig, M.T., Anderson, T.J.C., 2012. A Major Genome Region Underlying Artemisinin Resistance in Malaria. *Science* 336, 79-82.
- Cheeseman, I.H., McDew-White, M., Phyo, A.P., Sriprawat, K., Nosten, F., Anderson, T.J., 2015. Pooled sequencing and rare variant association tests for identifying the determinants of emerging drug resistance in malaria parasites. *Mol Biol Evol* 32, 1080-1090.
- Chen, H., Patterson, N., Reich, D., 2010. Population differentiation as a test for selective sweeps. *Genome Res* 20, 393-402.
- Chevalier, F.D., Le Clec'h, W., Eng, N., Rugel, A.R., Assis, R.R., Oliveira, G., Holloway, S.P., Cao, X., Hart, P.J., LoVerde, P.T., Anderson, T.J., 2016. Independent origins of loss-of-function mutations conferring oxamniquine resistance in a Brazilian schistosome population. *Int J Parasitol* 46, 417-424.

- Chevalier, F.D., Le Clec'h, W., McDew-White, M., Menon, V., Guzman, M.A., Holloway, S.P., Cao, X., Taylor, A.B., Kinung'hi, S., Gouvras, A.N., Webster, B.L., Webster, J.P., Emery, A.M., Rollinson, D., Garba Djirmay, A., Al Mashikhi, K.M., Al Yafae, S., Idris, M.A., Mone, H., Mouahid, G., Hart, P.J., LoVerde, P.T., Anderson, T.J.C., 2019. Oxamniquine resistance alleles are widespread in Old World *Schistosoma mansoni* and predate drug deployment. *PLoS Pathog* 15, e1007881.
- Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C., Ecker, J.R., Cantu, D., Rank, D.R., Schatz, M.C., 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13, 1050.
- Choi, Y.-J., Bisset, S.A., Doyle, S.R., Hallsworth-Pepin, K., Martin, J., Grant, W.N., Mitreva, M., 2017. Genomic introgression mapping of field-derived multiple-anthelmintic resistance in *Teladorsagia circumcincta*. *PLoS Genet* 13, e1006857.
- Christie, M., Jackson, F., 1982. Specific identification of strongyle eggs in small samples of sheep faeces. *Res Vet Sci* 32, 113-117.
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* 6, 80-92.
- Cole, V.G., 1986. Helminth parasites of sheep and cattle. Australian Government Publishing Service, Canberra.
- Coles, G., 2003. Strategies to minimize anthelmintic resistance in large animal practice. *In Practice* 25, 494-499.
- Coles, G.C., Bauer, C., Borgsteede, F.H.M., Geerts, S., Klei, T.R., Taylor, M.A., Waller, P.J., 1992. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.). Methods for the detection of anthelmintic resistance in nematodes of veterinary importance. *Vet Parasitol* 44.
- Coles, G.C., Jackson, F., Pomroy, W.E., Prichard, R.K., von Samson-Himmelstjerna, G., Silvestre, A., Taylor, M.A., Vercruyse, J., 2006. The detection of anthelmintic resistance in nematodes of veterinary importance. *Vet Parasitol* 136, 167-185.
- Collins, T., Millar, N.S., 2010. Nicotinic acetylcholine receptor transmembrane mutations convert ivermectin from a positive to a negative allosteric modulator. *Molecular pharmacology* 78, 198-204.
- Cook, A., Aptel, N., Portillo, V., Siney, E., Sihota, R., Holden-Dye, L., Wolstenholme, A., 2006. *Caenorhabditis elegans* ivermectin receptors regulate locomotor behaviour and are functional orthologues of *Haemonchus contortus* receptors. *Mol Biochem Parasitol* 147, 118-125.
- Coop, R.L., Angus, K.W., Sykes, A.R., 1979. Chronic infection with *Trichostrongylus vitrinus* in sheep. Pathological changes in the small intestine. *Res Vet Sci* 26, 363-371.
- Coop, R.L., Sykes, A.R., Spence, J.A., Aitchison, G.U., 1981. *Ostertagia circumcincta* infection of lambs, the effect of different intakes of larvae on skeletal development. *J Comp Pathol* 91, 521-530.
- Coop, R.L., Sykes, A.R., Angus, K.W., 1982. The effect of three levels of intake of *Ostertagia circumcincta* Larvae on growth rate, food intake and body composition of growing lambs. *J Agr Sci* 98, 247-255.
- Coop, R.L., Field, A.C., Graham, R.B., Angus, K.W., Jackson, F., 1986. Effect of concurrent infection with *Ostertagia circumcincta* and *Trichostrongylus vitrinus* on the performance of lambs. *Res Vet Sci* 40, 241-245.

- Cosmos, U.K. 2018. Daily and sub-daily hydrometeorological and soil data (2013-2016) [COSMOS-UK] v2 (NERC Environmental Information Data Centre).
- Coyne, M.J., Smith, G., Johnstone, C., 1991. Fecundity of gastrointestinal trichostrongylid nematodes of sheep in the field. *Am J Vet Res* 52, 1182-1188.
- Coyne, M.J., Smith, G., 1992. The mortality and fecundity of *Haemonchus contortus* in parasite-naive and parasite-exposed sheep following single experimental infections. *Int J Parasitol* 22, 315-325.
- Craig, B.H., Pilkington, J.G., Pemberton, J.M., 2006. Gastrointestinal nematode species burdens and host mortality in a feral sheep population. *Parasitol* 133, 485-496.
- Craig, B.H., Jones, O.R., Pilkington, J.G., Pemberton, J.M., 2009. Re-establishment of nematode infra-community and host survivorship in wild Soay sheep following anthelmintic treatment. *Vet Parasitol* 161, 47-52.
- Crilly, J.P., Sargison, N., 2015. Ruminant coprological examination: beyond the McMaster slide. In *Practice* 37, 68-76.
- Cringoli, G., Rinaldi, L., Veneziano, V., Capelli, G., Scala, A., 2004. The influence of flotation solution, sample dilution and the choice of McMaster slide area (volume) on the reliability of the McMaster technique in estimating the faecal egg counts of gastrointestinal strongyles and *Dicrocoelium dendriticum* in sheep. *Vet Parasitol* 123, 121-131.
- Crofton, H.D., Thomas, R.J., 1951. A New Species of Nematodirus in Sheep. *Nat* 168, 559-559.
- Crofton, H.D., 1965. Ecology and biological plasticity of sheep nematodes. I. The effect of temperature on the hatching of eggs of some nematode parasites of sheep. *Cornell Vet* 55, 242-250.
- Culetto, E., Baylis, H.A., Richmond, J.E., Jones, A.K., Fleming, J.T., Squire, M.D., Lewis, J.A., Sattelle, D.B., 2004. The *Caenorhabditis elegans unc-63* gene encodes a levamisole-sensitive nicotinic acetylcholine receptor alpha subunit. *J Biol Chem* 279, 42476-42483.
- Cully, D.F., Vassilatis, D.K., Liu, K.K., Paress, P.S., Van der Ploeg, L.H., Schaeffer, J.M., Arena, J.P., 1994. Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nat* 371, 707-711.
- Cutter, A.D., 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. *Genet* 172, 171-184.
- Cutter, A.D., Baird, S.E., Charlesworth, D., 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genet* 174, 901-913.
- Cvilink, V., Skálová, L., Szotáková, B., Lamka, J., Kostianen, R., Ketola, R.A., 2008. LC-MS-MS identification of albendazole and flubendazole metabolites formed *ex vivo* by *Haemonchus contortus*. *Anal Bioanal Chem* 391, 337-343.
- da Silveira, W.F., Braga, F.R., de Oliveira Tavela, A., dos Santos, L.F., Domingues, R.R., Aguiar, A.R., Ferraz, C.M., de Carvalho, L.M., de Hollanda Ayupe, T., Zanuncio, J.C., de Araújo, J.V., 2017. Nematophagous fungi combinations reduce free-living stages of sheep gastrointestinal nematodes in the field. *Journal Invertebr Pathol* 150, 1-5.
- Dabney, J., Meyer, M., 2012. Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTechniques* 52, 87-94.

- Daborn, P.J., Yen, J.L., Bogwitz, M.R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., Feyereisen, R., Wilson, T.G., ffrench-Constant, R.H., 2002. A Single P450 Allele Associated with Insecticide Resistance in *Drosophila*. *Science* 297, 2253-2256.
- DaCosta, J.M., Sorenson, M.D., 2014. Amplification Biases and Consistent Recovery of Loci in a Double-Digest RAD-seq Protocol. *PLoS One* 9, e106713.
- DaCosta, J.M., Sorenson, M.D., 2016. ddRAD-seq phylogenetics based on nucleotide, indel, and presence–absence polymorphisms: Analyses of two avian genera with contrasting histories. *Mol Phylogenet Evol* 94, 122-135.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis, G., 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156-2158.
- Dash, K.M., Hall, E., Barger, I.A., 1988. The role of arithmetic and geometric mean worm egg counts in faecal egg count reduction tests and in monitoring strategic drenching programs in sheep. *Aust Vet J* 65, 66-68.
- de Lourdes Mottier, M., Prichard, R.K., 2008. Genetic analysis of a relationship between macrocyclic lactone and benzimidazole anthelmintic selection on *Haemonchus contortus*. *Pharmacogenet Genomics* 18, 129-140.
- Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M., Lasken, R.S., 2002. Comprehensive human genome amplification using multiple displacement amplification. *P Natl A Sci* 99, 5261-5266.
- DEFRA 2018. Livestock numbers in the UK.
- Demeler, J., Küttler, U., El-Abdellati, A., Stafford, K., Rydzik, A., Varady, M., Kenyon, F., Coles, G., Höglund, J., Jackson, F., 2010a. Standardization of the larval migration inhibition test for the detection of resistance to ivermectin in gastro intestinal nematodes of ruminants. *Vet Parasitol* 174.
- Demeler, J., Küttler, U., von Samson-Himmelstjerna, G., 2010b. Adaptation and evaluation of three different *in vitro* tests for the detection of resistance to anthelmintics in gastro intestinal nematodes of cattle. *Vet Parasitol* 170.
- Dent, J.A., Davis, M.W., Avery, L., 1997. *avr-15* encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in *Caenorhabditis elegans*. *EMBO J* 16, 5867-5879.
- Dent, J.A., Smith, M.H.M., Vassilatis, D.K., Avery, L., 2000. The genetics of ivermectin resistance in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 97, 2674-2679.
- Deus, K.M., Saavedra-Rodriguez, K., Butters, M.P., Black, W.C.t., Foy, B.D., 2012. The effect of ivermectin in seven strains of *Aedes aegypti* (Diptera: *Culicidae*) including a genetically diverse laboratory strain and three permethrin resistant strains. *J Med Entomol* 49, 356-363.
- Dever, M.L., Kahn, L.P., 2015. Decline in faecal worm egg counts in lambs suckling ewes treated with lipophilic anthelmintics: implications for hastening development of anthelmintic resistance. *Vet Parasitol* 209, 229-234.
- Dicker, A.J., Nath, M., Yaga, R., Nisbet, A.J., Lainson, F.A., Gilleard, J.S., Skuce, P.J., 2011a. *Teladorsagia circumcincta*: the transcriptomic response of a multi-drug-resistant isolate to ivermectin exposure *in vitro*. *Exp Parasitol* 127, 351-356.

- Dicker, A.J., Nisbet, A.J., Skuce, P.J., 2011b. Gene expression changes in a P-glycoprotein (*Tci-pgp-9*) putatively associated with ivermectin resistance in *Teladorsagia circumcincta*. *Int J Parasitol* 41, 935-942.
- Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D., McMichael, J.F., Wallis, J.W., Lu, C., Shen, D., Harris, C.C., Dooling, D.J., Fulton, R.S., Fulton, L.L., Chen, K., Schmidt, H., Kalicki-Veizer, J., Magrini, V.J., Cook, L., McGrath, S.D., Vickery, T.L., Wendl, M.C., Heath, S., Watson, M.A., Link, D.C., Tomasson, M.H., Shannon, W.D., Payton, J.E., Kulkarni, S., Westervelt, P., Walter, M.J., Graubert, T.A., Mardis, E.R., Wilson, R.K., DiPersio, J.F., 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nat* 481, 506-510.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.
- Dobson, R.J., Donald, A.D., Waller, P.J., Snowdon, K.L., 1986. An egg-hatch assay for resistance to levamisole in trichostrongyloid nematode parasites. *Vet Parasitol* 19, 77-84.
- Dobson, R.J., Barnes, E.H., Bircilijn, S.D., Gill, J.H., 1992. The survival of *Ostertagia circumcincta* and *Trichostrongylus colubriformis* in faecal culture as a source of bias in apportioning egg counts to worm species. *Int J Parasitol* 22.
- Dogovski, C., Xie, S.C., Burgio, G., Bridgford, J., Mok, S., McCaw, J.M., Chotivanich, K., Kenny, S., Gnadig, N., Straimer, J., Bozdech, Z., Fidock, D.A., Simpson, J.A., Dondorp, A.M., Foote, S., Klonis, N., Tilley, L., 2015. Targeting the cell stress response of *Plasmodium falciparum* to overcome artemisinin resistance. *PLoS Biol* 13, e1002132.
- Doherty, E., Burgess, S., Mitchell, S., Wall, R., 2018. First evidence of resistance to macrocyclic lactones in *Psoroptes ovis* sheep scab mites in the UK. *Vet Rec* 182, 106-106.
- Dou, Q., Chen, H.N., Wang, K., Yuan, K., Lei, Y., Li, K., Lan, J., Chen, Y., Huang, Z., Xie, N., Zhang, L., Xiang, R., Nice, E.C., Wei, Y., Huang, C., 2016. Ivermectin Induces Cytostatic Autophagy by Blocking the PAK1/Akt Axis in Breast Cancer. *Cancer Res* 76, 4457-4469.
- Doyle, S.R., Bourguinat, C., Nana-Djeunga, H.C., Kengne-Ouafo, J.A., Pion, S.D.S., Bopda, J., Kamgno, J., Wanji, S., Che, H., Kuesel, A.C., Walker, M., Basanez, M.G., Boakye, D.A., Osei-Atweneboana, M.Y., Boussinesq, M., Prichard, R.K., Grant, W.N., 2017. Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. *PLoS Negl Trop Dis* 11, e0005816.
- Doyle, S.R., Laing, R., Bartley, D.J., Britton, C., Chaudhry, U., Gilleard, J.S., Holroyd, N., Mable, B.K., Maitland, K., Morrison, A.A., Tait, A., Tracey, A., Berriman, M., Devaney, E., Cotton, J.A., Sargison, N.D., 2018. A Genome Resequencing-Based Genetic Map Reveals the Recombination Landscape of an Outbred Parasitic Nematode in the Presence of Polyploidy and Polyandry. *Genome Biol Evol* 10, 396-409.
- Doyle, S.R., Illingworth, C.J.R., Laing, R., Bartley, D.J., Redman, E., Martinelli, A., Holroyd, N., Morrison, A.A., Rezansoff, A., Tracey, A., Devaney, E., Berriman, M., Sargison, N., Cotton, J.A., Gilleard, J.S., 2019. Population genomic and evolutionary modelling analyses reveal a single major QTL for

- ivermectin drug resistance in the pathogenic nematode, *Haemonchus contortus*. BMC Genomics 20, 218.
- Driscoll, M., Dean, E., Reilly, E., Bergholz, E., Chalfie, M., 1989. Genetic and molecular analysis of a *Caenorhabditis elegans* beta-tubulin that conveys benzimidazole sensitivity. J Cell Biol 109, 2993-3003.
- Drudge, J.H., Szanto, J., Wyant, Z.N., Elam, G., 1964. Field Studies on Parasite Control in Sheep: Comparison of Thiabensazole, Ruelene, and Phenothiazine. Am J Vet Res 25, 1512-1518.
- Duan, Y., Yang, Y., Li, T., Zhao, D., Cao, J., Shi, Y., Wang, J., Zhou, M., 2016. Development of a rapid and high-throughput molecular method for detecting the F200Y mutant genotype in benzimidazole-resistant isolates of *Fusarium asiaticum*. Pest Manag Sci 72, 2128-2135.
- Duan, Y.B., Yang, Y., Wang, J.X., Chen, C.J., Steinberg, G., Fraaije, B.A., Zhou, M.G., 2018. Simultaneous Detection of Multiple Benzimidazole-Resistant beta-Tubulin Variants of *Botrytis cinerea* using Loop-Mediated Isothermal Amplification. Plant Dis 102, 2016-2024.
- Duke, B.O., Zea-Flores, G., Munoz, B., 1991. The embryogenesis of *Onchocerca volvulus* over the first year after a single dose of ivermectin. Trop Med Parasitol 42, 175-180.
- Durham, P.J., Elliott, D.C., 1975. Experimental *Ostertagia* infection of sheep: pathology following a single high dose of larvae. N Z Vet J 23, 193-196.
- Ehrenreich, I.M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J.A., Gresham, D., Caudy, A.A., Kruglyak, L., 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nat 464, 1039-1042.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korfach, J., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science 323, 133-138.
- El-Abdellati, A., De Graef, J., Van Zeveren, A., Donnan, A., Skuce, P., Walsh, T., Wolstenholme, A., Tait, A., Vercruyse, J., Claerebout, E., Geldhof, P., 2011. Altered *avr-14B* gene transcription patterns in ivermectin-resistant isolates of the cattle parasites, *Cooperia oncophora* and *Ostertagia ostertagi*. Int J Parasitol 41, 951-957.
- Elard, L., Sauve, C., Humbert, J.F., 1998. Fitness of benzimidazole-resistant and -susceptible worms of *Teladorsagia circumcincta*, a nematode parasite of small ruminants. Parasitol 117 ( Pt 6), 571-578.
- Elard, L., Humbert, J.F., 1999. Importance of the mutation of amino acid 200 of the isotype 1  $\beta$ -tubulin gene in the benzimidazole resistance of the small-ruminant parasite *Teladorsagia circumcincta*. Parasitol Res 85, 452-456.
- Engels, W.R., 2009. Exact Tests for Hardy–Weinberg Proportions. Genet 183, 1431-1441.
- Esteban-Ballesteros, M., Rojo-Vázquez, F.A., Skuce, P.J., Melville, L., González-Lanza, C., Martínez-Valladares, M., 2017. Quantification of resistant alleles in the  $\beta$ -tubulin gene of field strains of gastrointestinal nematodes and their relation with the faecal egg count reduction test. BMC Vet Res 13, 71-71.



- Ewels, P., Magnusson, M., Lundin, S., Källér, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047-3048.
- Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1, 47-50.
- Fairlie-Clarke, K., Kaseja, K., Sotomaior, C., Brady, N., Moore, K., Stear, M., 2019. Salivary IgA: A biomarker for resistance to *Teladorsagia circumcincta* and a new estimated breeding value. *Vet Parasitol* 269, 16-20.
- Fauvin, A., Charvet, C., Issouf, M., Cortet, J., Cabaret, J., Neveu, C., 2010. cDNA-AFLP analysis in levamisole-resistant *Haemonchus contortus* reveals alternative splicing in a nicotinic acetylcholine receptor subunit. *Mol Biochem Parasitol* 170, 105-107.
- Fenker, K.E., Hansen, A.A., Chong, C.A., Jud, M.C., Duffy, B.A., Paul Norton, J., Hansen, J.M., Stanfield, G.M., 2014. SLC6 family transporter SNF-10 is required for protease-mediated activation of sperm motility in *C. elegans*. *Dev Biol* 393, 171-182.
- Fitzpatrick, J.L., 2013. Global food security: the impact of veterinary parasites and parasitologists. *Vet Parasitol* 195, 233-248.
- Fleming, J.T., Squire, M.D., Barnes, T.M., Tornoe, C., Matsuda, K., Ahnn, J., Fire, A., Sulston, J.E., Barnard, E.A., Sattelle, D.B., Lewis, J.A., 1997. *Caenorhabditis elegans* levamisole resistance genes *lev-1*, *unc-29*, and *unc-38* encode functional nicotinic acetylcholine receptor subunits. *J Neurosci* 17, 5843-5857.
- Fleming, J.T., Baylis, H.A., Sattelle, D.B., Lewis, J.A., 2009. Molecular cloning and *in vitro* expression of *C. elegans* and parasitic nematode ionotropic receptors. *Parasitol* 113, S175-S190.
- Fontaine, P., Choe, K., 2018. The transcription factor SKN-1 and detoxification gene *ugt-22* alter albendazole efficacy in *Caenorhabditis elegans*. *Int J Parasitol Drugs Drug Resist* 8, 312-319.
- Franks, D.M., Izumikawa, T., Kitagawa, H., Sugahara, K., Okkema, P.G., 2006. *C. elegans* pharyngeal morphogenesis requires both de novo synthesis of pyrimidines and synthesis of heparan sulfate proteoglycans. *Dev Biol* 296, 409-420.
- Fraser, D.E., Hunt, P.J., Skinner, R.J., Coles, G.C., 2006. Survey of parasite control on sheep farms in south-west England. *Vet Rec* 158, 55-57.
- Futschik, A., Schlotterer, C., 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genet* 186, 207-218.
- Gabriel, S., Ziaugra, L., Tabbaa, D., 2009. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet* Chapter 2, Unit 2.12.
- Gallardo, F., Mariame, B., Gence, R., Tilkin-Mariame, A.F., 2018. Macrocyclic lactones inhibit nasopharyngeal carcinoma cells proliferation through PAK1 inhibition and reduce *in vivo* tumor growth. *Drug Des Devel Ther* 12, 2805-2814.
- Gautier, M., Foucaud, J., Gharbi, K., Cezard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhue, C., Estoup, A., 2013a. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol* 22, 3766-3779.
- Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhue, C., Pudlo, P., Cornuet, J.M., Estoup, A., 2013b. The effect of RAD allele dropout on the

- estimation of genetic variation within and between populations. *Mol Ecol* 22, 3165-3178.
- Geerts, S., Brandt, J., Borgsteede, F.H., Van Loon, H., 1989. Reliability and reproducibility of the larval paralysis test as an *in vitro* method for the detection of anthelmintic resistance of nematodes against levamisole and morantel tartrate. *Vet Parasitol* 30, 223-232.
- Gendrel, M., Rapti, G., Richmond, J.E., Bessereau, J.-L., 2009. A secreted complement-control-related protein ensures acetylcholine receptor clustering. *Nat* 461, 992.
- George, M.M., Lopez-Soberal, L., Storey, B.E., Howell, S.B., Kaplan, R.M., 2017. Motility in the L3 stage is a poor phenotype for detecting and measuring resistance to avermectin/milbemycin drugs in gastrointestinal nematodes of livestock. *Int J Parasitol Drugs Drug Resist* 8, 22-30.
- Geurden, T., Hodge, A., Noe, L., Winstanley, D., Bartley, D.J., Taylor, M., Morgan, C., Fraser, S.J., Maeder, S., Bartram, D., 2012. The efficacy of a combined oral formulation of derquantel-abamectin against anthelmintic resistant gastro-intestinal nematodes of sheep in the UK. *Vet Parasitol* 189, 308-316.
- Ghisi, M., Kaminsky, R., Mäser, P., 2007. Phenotyping and genotyping of *Haemonchus contortus* isolates reveals a new putative candidate mutation for benzimidazole resistance in nematodes. *Vet Parasitol* 144, 313-320.
- Gibbons, L.M., Khalil, L.F., 1982. A key for the identification of genera of the nematode family Trichostrongylidae Leiper, 1912. *J Helminthol* 56, 185-233.
- Gibbs, H.C., 1986a. Hypobiosis in parasitic nematodes—an update. *Adv Parasitol* 25.
- Gibbs, H.C., 1986b. Hypobiosis and the periparturient rise in sheep. *Vet Clin North Am Food Anim Pract* 2.
- Gibson, T.E., Everett, G., 1971. The seasonal fluctuations of the larval populations of four species of trichostrongylid nematode on pasture herbage. *Res Vet Sci* 12, 602-604.
- Gibson, T.E., Everett, G., 1972. The ecology of the free-living stages of *Ostertagia circumcincta*. *Parasitol* 64, 451-460.
- Gibson, T.E., 1973. Recent Advances in Epidemiology and Control of Parasitic Gastroenteritis in Sheep. *Vet Rec* 92, 469-473.
- Gibson, T.E., Parfitt, J.W., 1977. Egg output of *Ostertagia circumcincta* in sheep given single infections of varying size. *Vet Parasitol* 3, 61-66.
- Gill, J.H., Redwin, J.M., Van Wyk, J.A., Lacey, E., 1995. Avermectin inhibition of larval development in *Haemonchus contortus* — Effects of ivermectin resistance. *Int J Parasitol* 25, 463-470.
- Giudici, C., Aumont, G., Mahieu, M., Saulai, M., Cabaret, J., 1999. Changes in gastro-intestinal helminth species diversity in lambs under mixed grazing on irrigated pastures in the tropics (French West Indies). *Vet Res* 30, 573-581.
- Glaubitz, J.C., 2004. CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol Ecol Notes* 4, 309-310.
- Glendinning, S.K., Buckingham, S.D., Sattelle, D.B., Wonnacott, S., Wolstenholme, A.J., 2011. Glutamate-Gated Chloride Channels of *Haemonchus contortus* Restore Drug Sensitivity to Ivermectin Resistant *Caenorhabditis elegans*. *PLoS One* 6.
- Glover, M., Clarke, C., Nabb, L., Schmidt, J., 2017. Anthelmintic efficacy on sheep farms in south-west England. *Vet Rec* 180, 378-378.

- Godber, O.F., Phythian, C.J., Bosco, A., Ianniello, D., Coles, G., Rinaldi, L., Cringoli, G., 2015. A comparison of the FECPAK and Mini-FLOTAC faecal egg counting techniques. *Vet Parasitol* 207, 342-345.
- Godoy, P., Che, H., Beech, R.N., Prichard, R.K., 2015. Characterization of *Haemonchus contortus* P-glycoprotein-16 and its interaction with the macrocyclic lactone anthelmintics. *Mol Biochem Parasitol* 204, 11-15.
- Godoy, P., Che, H., Beech, R.N., Prichard, R.K., 2016. Characterisation of P-glycoprotein-9.1 in *Haemonchus contortus*. *Parasit Vectors* 9, 52.
- Gonzalez, J.F., Hernandez, A., Molina, J.M., Fernandez, A., Raadsma, H.W., Meeusen, E.N., Piedrafita, D., 2008. Comparative experimental *Haemonchus contortus* infection of two sheep breeds native to the Canary Islands. *Vet Parasitol* 153, 374-378.
- Good, B., Hanrahan, J.P., de Waal, D.T., Patten, T., Kinsella, A., Lynch, C.O., 2012. Anthelmintic-resistant nematodes in Irish commercial sheep flocks—the state of play. *Ir Vet J* 65, 21.
- Gordon, H.M., 1961. Thiabendazole: a Highly Effective Anthelmintic for Sheep. *Nat* 191, 1409-1410.
- Grillo, V., Jackson, F., Gilleard, J.S., 2006. Characterisation of *Teladorsagia circumcincta* microsatellites and their development as population genetic markers. *Mol Biochem Parasitol* 148, 181-189.
- Grillo, V., Jackson, F., Cabaret, J., Gilleard, J.S., 2007. Population genetic analysis of the ovine parasitic nematode *Teladorsagia circumcincta* and evidence for a cryptic species. *Int J Parasitol* 37, 435-447.
- Grillo, V.L., 2005. Development of microsatellites and the population genetic analysis of the parasitic nematode *Teladorsagia circumcincta*. PhD. University of Glasgow, Glasgow.
- Grimshaw, W.T., Hunt, K.R., Hong, C., Coles, G.C., 1994. Detection of anthelmintic resistant nematodes in sheep in southern England by a faecal egg count reduction test. *Vet Rec* 135, 372-374.
- Guo, S.W., Thompson, E.A., 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48, 361-372.
- Hahnel, S.R., Zdraljevic, S., Rodriguez, B.C., Zhao, Y., McGrath, P.T., Andersen, E.C., 2018. Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog* 14, e1007226.
- Hamer, K., Bartley, D., Jennings, A., Morrison, A., Sargison, N., 2018. Lack of efficacy of monepantel against trichostrongyle nematodes in a UK sheep flock. *Vet Parasitol* 257, 48-53.
- Hamer, K., McIntyre, J., Morrison, A.A., Jennings, A., Kelly, R.F., Leeson, S., Bartley, D.J., Chaudhry, U., Busin, V., Sargison, N., 2019. The dynamics of ovine gastrointestinal nematode infections within ewe and lamb cohorts on three Scottish sheep farms. *Prev Vet Med* 171, 104752.
- Harris, D.T., Horvitz, H.R., 2011. MAB-10/NAB acts with LIN-29/EGR to regulate terminal differentiation and the transition from larva to adult in *C. elegans*. *Development* 138, 4051-4062.
- Hartl, D.L., Clark, A.G., 2007. Principles of population genetics, Fourth edition. Edition. Sinauer Associates, Sunderland, Mass., xv, 652 pages pp.
- Heim, C., Hertzberg, H., Butschi, A., Bleuler-Martinez, S., Aebi, M., Deplazes, P., Künzler, M., Štefanić, S., 2015. Inhibition of *Haemonchus contortus* larval development by fungal lectins. *Parasit Vectors* 8, 425.
- Hernando, G., Bouzat, C., 2014. *Caenorhabditis elegans* neuromuscular junction: GABA receptors and ivermectin action. *PLoS One* 9, e95072.

- Hodgkin, J., 1983. Male Phenotypes and Mating Efficiency in *Caenorhabditis elegans*. *Genet* 103, 43-64.
- Hodgkinson, J., Cwiklinski, K., Beesley, N.J., Paterson, S., Williams, D.J., 2013. Identification of putative markers of triclabendazole resistance by a genome-wide analysis of genetically recombinant *Fasciola hepatica*. *Parasitol* 140, 1523-1533.
- Hodgkinson, J.E., Cwiklinski, K., Beesley, N., Hartley, C., Allen, K., Williams, D.J.L., 2018. Clonal amplification of *Fasciola hepatica* in *Galba truncatula*: within and between isolate variation of triclabendazole-susceptible and -resistant clones. *Parasit Vectors* 11, 363.
- Hodgkinson, J.E., Kaplan, R.M., Kenyon, F., Morgan, E.R., Park, A.W., Paterson, S., Babayan, S.A., Beesley, N.J., Britton, C., Chaudhry, U., Doyle, S.R., Ezenwa, V.O., Fenton, A., Howell, S.B., Laing, R., Mable, B.K., Matthews, L., McIntyre, J., Milne, C.E., Morrison, T.A., Prentice, J.C., Sargison, N.D., Williams, D.J.L., Wolstenholme, A.J., Devaney, E., 2019. Refugia and anthelmintic resistance: Concepts and challenges. *Int J Parasitol Drugs Drug Resist* 10, 51-57.
- Hoekstra, R., Otsen, M., Tibben, J., Lenstra, J.A., Roos, M.H., 2000. Transposon associated markers for the parasitic nematode *Haemonchus contortus*. *Mol Biochem Parasitol* 105, 127-135.
- Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., Stanke, M., 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32, 767-769.
- Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6, e1000862.
- Hohenlohe, P.A., Catchen, J., Cresko, W.A., 2012. Population Genomic Analysis of Model and Nonmodel Organisms Using Sequenced RAD Tags, In: Pompanon, F., Bonin, A. (Eds.) *Data Production and Analysis in Population Genomics: Methods and Protocols*. Humana Press, Totowa, NJ, pp. 235-260.
- Holden-Dye, L., Walker, R.J., 1990. Avermectin and avermectin derivatives are antagonists at the 4-aminobutyric acid (GABA) receptor on the somatic muscle cells of *Ascaris*; is this the site of anthelmintic action? *Parasitol* 101 Pt 2, 265-271.
- Holden-Dye, L., Joyner, M., O'Connor, V., Walker, R.J., 2013. Nicotinic acetylcholine receptors: a comparison of the nAChRs of *Caenorhabditis elegans* and parasitic nematodes. *Parasitol Int* 62, 606-615.
- Hong, C., Hunt, K.R., Coles, G.C., 1994. Resistance to levamisole. *Vet Rec* 135, 535-536.
- Hong, C., Hunt, K.R., Coles, G.C., 1996. Occurrence of anthelmintic resistant nematodes on sheep farms in England and goat farms in England and Wales. *Vet Rec* 139, 83-86.
- Horoszok, L., Raymond, V., Sattelle, D.B., Wolstenholme, A.J., 2001. GLC-3: a novel fipronil and BIDN-sensitive, but picrotoxinin-insensitive, L-glutamate-gated chloride channel subunit from *Caenorhabditis elegans*. *Br J Pharmacol* 132, 1247-1254.
- Hosking, B.C., Kaminsky, R., Sager, H., Rolfe, P.F., Seewald, W., 2009. A pooled analysis of the efficacy of monepantel, an amino-acetonitrile derivative against gastrointestinal nematodes of sheep. *Parasitol Res* 106, 529.
- Houdijk, J.G.M., Kyriazakis, I., Jackson, F., Huntley, J.F., Coop, R.L., 2000. Can an increased intake of metabolizable protein affect the periparturient

- relaxation in immunity against *Teladorsagia circumcincta* in sheep? *Vet Parasitol* 91, 43-62.
- Howe, K.L., Bolt, B.J., Cain, S., Chan, J., Chen, W.J., Davis, P., Done, J., Down, T., Gao, S., Grove, C., Harris, T.W., Kishore, R., Lee, R., Lomax, J., Li, Y., Muller, H.M., Nakamura, C., Nuin, P., Paulini, M., Raciti, D., Schindelman, G., Stanley, E., Tuli, M.A., Van Auken, K., Wang, D., Wang, X., Williams, G., Wright, A., Yook, K., Berriman, M., Kersey, P., Schedl, T., Stein, L., Sternberg, P.W., 2016. WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res* 44, D774-780.
- Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., Berriman, M., 2017. WormBase ParaSite - a comprehensive resource for helminth genomics. *Mol Biochem Parasitol* 215, 2-10.
- Hubert, J., Kerboeuf, D., 1992. A microlarval development assay for the detection of anthelmintic resistance in sheep nematodes. *Vet Rec* 130, 442-446.
- Huynh, J.M., Dang, H., Munoz-Tucker, I.A., O'Ketch, M., Liu, I.T., Perno, S., Bhuyan, N., Crain, A., Borbon, I., Fares, H., 2016. ESCRT-Dependent Cell Death in a *Caenorhabditis elegans* Model of the Lysosomal Storage Disorder Mucopolysaccharidosis Type IV. *Genet* 202, 619-638.
- Hwang, J.M., Chang, D.J., Kim, U.S., Lee, Y.S., Park, Y.S., Kaang, B.K., Cho, N.J., 1999. Cloning and functional characterization of a *Caenorhabditis elegans* muscarinic acetylcholine receptor. *Receptor Channel* 6, 415-424.
- Hybu Cig Cymru, H., 2015. Wales Against Anthelmintic Resistance Development, Final Project Report.
- Illumina 2011. Quality Scores for Next-Generation Sequencing Assessing sequencing accuracy using Phred quality scoring.
- Israf, D.A., Jackson, F., Stevenson, L.M., Jones, D.G., Jackson, J.E., Huntley, J.F., Coop, R.L., 1997. Persistence of immunity of *Nematodirus battus* infection in lambs. *Vet Parasitol* 71, 39-52.
- Jack, C., Hotchkiss, E., Sargison, N.D., Toma, L., Milne, C., Bartley, D.J., 2017. A quantitative analysis of attitudes and behaviours concerning sustainable parasite control practices from Scottish sheep farmers. *Prev Vet Med* 139, 134-145.
- Jackson, F., Jackson, E., Williams, J.T., 1988. Susceptibility of the pre-parturient ewe to infection with *Trichostrongylus vitrinus* and *Ostertagia circumcincta*. *Res Vet Sci* 45, 213-218.
- Jackson, F., Jackson, E., Coop, R.L., 1992a. Evidence of multiple anthelmintic resistance in a strain of *Teladorsagia circumcincta* (*Ostertagia circumcincta*) isolated from goats in Scotland. *Res Vet Sci* 53, 371-374.
- Jackson, F., Jackson, E., Coop, R.L., Huntley, J., 1992b. Interactions between *Teladorsagia circumcincta* and *Trichostrongylus vitrinus* infections in young lambs. *Res Vet Sci* 53, 363-370.
- James, C.E., Davey, M.W., 2009. Increased expression of ABC transport proteins is associated with ivermectin resistance in the model nematode *Caenorhabditis elegans*. *Int J Parasitol* 39, 213-220.
- Jansen, J., 1987. The peri-parturient rise in sheep. Faecal worm egg counts in normal and late lambing ewes. *Vet Q* 9, 97-102.
- Janssen, I.J., Krucken, J., Demeler, J., von Samson-Himmelstjerna, G., 2015. Transgenically expressed *Parascaris* P-glycoprotein-11 can modulate ivermectin susceptibility in *Caenorhabditis elegans*. *Int J Parasitol Drugs Drug Resist* 5, 44-47.

- Jarrett, W.F., Jennings, F.W., Mc, I.W., Mulligan, W., Urquhart, G.M., 1960. Immunological studies on *Dictyocaulus viviparus* infection; immunity produced by the administration of irradiated larvae. *Immunology* 3, 145-151.
- Jeffcoate, I.A., Wedrychowicz, H., Fishwick, G., Dunlop, E.M., Duncan, J.L., Holmes, P.H., 1992. Pathophysiology of the periparturient egg rise in sheep: a possible role for IgA. *Res Vet Sci* 53, 212-218.
- Jin, L., Feng, X., Rong, H., Pan, Z., Inaba, Y., Qiu, L., Zheng, W., Lin, S., Wang, R., Wang, Z., Wang, S., Liu, H., Li, S., Xie, W., Li, Y., 2013. The antiparasitic drug ivermectin is a novel FXR ligand that regulates metabolism. *Nat Commun* 4, 1937.
- Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403-1405.
- Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070-3071.
- Jordan, H.E., Phillips, W.A., Morrison, R.D., Doyle, J.J., McKenzie, K., 1988. A 3-year study of continuous mixed grazing of cattle and sheep: Parasitism of offspring. *Int J Parasitol* 18, 779-784.
- Juliano, R.L., Ling, V., 1976. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim Biophys Acta* 455, 152-162.
- Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S.S., Wenger, A., Wieland-Berghausen, S., Goebel, T., 2008. A new class of anthelmintics effective against drug-resistant nematodes. *Nat* 452.
- Kaminsky, R., Mosimann, D., Sager, H., Stein, P., Hosking, B., 2009. Determination of the effective dose rate for monepantel (AAD 1566) against adult gastro-intestinal nematodes in sheep. *Int J Parasitol* 39, 443-446.
- Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281.
- Kamvar, Z.N., Brooks, J.C., Grünwald, N.J., 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet* 6.
- Kaplan, R.M., 2004. Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol* 20.
- Kaplan, R.M., Vidyashankar, A.N., 2012. An inconvenient truth: global worming and anthelmintic resistance. *Vet Parasitol* 186, 70-78.
- Keane, O.M., Keegan, J.D., Good, B., de Waal, T., Fanning, J., Gottstein, M., Casey, M., Hurley, C., Sheehan, M., 2014. High level of treatment failure with commonly used anthelmintics on Irish sheep farms. *Ir Vet J* 67, 16.
- Keegan, J.D., Good, B., de Waal, T., Fanning, J., Keane, O.M., 2017a. Genetic basis of benzimidazole resistance in *Teladorsagia circumcincta* in Ireland. *Ir Vet J* 70, 8.
- Keegan, J.D., Keane, O.M., Good, B., De Waal, T., Denny, M., Hanrahan, J.P., Fitzgerald, W., Sheehan, M., 2017b. A nationwide survey of anthelmintic treatment failure on sheep farms in Ireland. *Ir Vet J* 70, 7.
- Kelley, J.M., Elliott, T.P., Beddoe, T., Anderson, G., Skuce, P., Spithill, T.W., 2016. Current Threat of Triclabendazole Resistance in *Fasciola hepatica*. *Trends Parasitol* 32, 458-469.
- Kenyon, F., Greer, A.W., Coles, G.C., Cringoli, G., Papadopoulos, E., Cabaret, J., Berrag, B., Varady, M., Wyk, J.A., Thomas, E., Vercruyssen, J., Jackson, F., 2009a. The role of targeted selective treatments in the development of

- refugia-based approaches to the control of gastrointestinal nematodes of small ruminants. *Vet Parasitol* 164, 3-11.
- Kenyon, F., Sargison, N.D., Skuce, P.J., Jackson, F., 2009b. Sheep helminth parasitic disease in south eastern Scotland arising as a possible consequence of climate change. *Vet Parasitol* 163, 293-297.
- Kenyon, F., McBean, D., Greer, A.W., Burgess, C.G., Morrison, A.A., Bartley, D.J., Bartley, Y., Devin, L., Nath, M., Jackson, F., 2013. A comparative study of the effects of four treatment regimes on ivermectin efficacy, body weight and pasture contamination in lambs naturally infected with gastrointestinal nematodes in Scotland. *Int J Parasitol Drugs Drug Resist* 3, 77-84.
- Kenyon, F., Rinaldi, L., McBean, D., Pepe, P., Bosco, A., Melville, L., Devin, L., Mitchell, G., Ianniello, D., Charlier, J., Vercruyse, J., Cringoli, G., Levecke, B., 2016. Pooling sheep faecal samples for the assessment of anthelmintic drug efficacy using McMaster and Mini-FLOTAC in gastrointestinal strongyle and *Nematodirus* infection. *Vet Parasitol*.
- Kerboeuf, D., Hubert, J., Cardinaud, B., Blond-Riou, F., 1995. The persistence of the efficacy of injectable or oral moxidectin against *Teladorsagia*, *Haemonchus* and *Trichostrongylus* species in experimentally infected sheep. *Vet Rec* 137, 399.
- Kibbe, W.A., 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35, W43-46.
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12, 357.
- Kingan, S.B., Heaton, H., Cudini, J., Lambert, C.C., Baybayan, P., Galvin, B.D., Durbin, R., Korfach, J., Lawniczak, M.K.N., 2019. A High-Quality *De novo* Genome Assembly from a Single Mosquito Using PacBio Sequencing. *Genes* 10, 62.
- Knaus, B.J., Grünwald, N.J., 2017. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17, 44-53.
- Knox, M.R., 2002. Effectiveness of copper oxide wire particles for *Haemonchus contortus* control in sheep. *Aust Vet J* 80, 224-227.
- Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., Schlotterer, C., 2011a. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS One* 6, e15925.
- Kofler, R., Pandey, R.V., Schlotterer, C., 2011b. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27, 3435-3436.
- Kofler, R., Gomez-Sanchez, D., Schlotterer, C., 2016a. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Mol Biol Evol* 33, 2759-2764.
- Kofler, R., Nolte, V., Schlotterer, C., 2016b. The impact of library preparation protocols on the consistency of allele frequency estimates in Pool-Seq data. *Mol Ecol Resour* 16, 118-122.
- Kotze, A.C., Cowling, K., Bagnall, N.H., Hines, B.M., Ruffell, A.P., Hunt, P.W., Coleman, G.T., 2012. Relative level of thiabendazole resistance associated with the E198A and F200Y SNPs in larvae of a multi-drug resistant isolate of *Haemonchus contortus*. *Int J Parasitol Drugs Drug Resist* 2, 92-97.
- Kotze, A.C., Hunt, P.W., Skuce, P., von Samson-Himmelstjerna, G., Martin, R.J., Sager, H., Krucken, J., Hodgkinson, J., Lespine, A., Jex, A.R., Gilleard, J.S., Beech, R.N., Wolstenholme, A.J., Demeler, J., Robertson, A.P., Charvet, C.L., Neveu, C., Kaminsky, R., Rufener, L., Alberich, M., Menez,

- C., Prichard, R.K., 2014. Recent advances in candidate-gene and whole-genome approaches to the discovery of anthelmintic resistance markers and the description of drug/receptor interactions. *Int J Parasitol Drugs Drug Resist* 4, 164-184.
- Krause, R.M., Buisson, B., Bertrand, S., Corringer, P.-J., Galzi, J.-L., Changeux, J.-P., Bertrand, D., 1998. Ivermectin: A Positive Allosteric Effector of the  $\alpha 7$  Neuronal Nicotinic Acetylcholine Receptor. *Molecular pharmacology* 53, 283-294.
- Kuznetsova, A.Y., Seget, K., Moeller, G.K., de Pagter, M.S., de Roos, J.A., Durrbaum, M., Kuffer, C., Muller, S., Zaman, G.J., Kloosterman, W.P., Storchova, Z., 2015. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. *Cell Cycle* 14, 2810-2820.
- Kwa, M.S., Veenstra, J.G., Van Dijk, M., Roos, M.H., 1995. Beta-tubulin genes from the parasitic nematode *Haemonchus contortus* modulate drug resistance in *Caenorhabditis elegans*. *J Mol Biol* 246, 500-510.
- Kwa, M.S.G., Kooyman, F.N.J., Boersema, J.H., Roos, M.H., 1993. Effect of Selection for Benzimidazole Resistance in *Haemonchus contortus* on  $\beta$ -Tubulin Isotype 1 and Isotype 2 Genes. *Biochem Biophys Res Commun* 191, 413-419.
- Kwa, M.S.G., Veenstra, J.G., Roos, M.H., 1994. Benzimidazole resistance in *Haemonchus contortus* is correlated with a conserved mutation at amino acid 200 in  $\beta$ -tubulin isotype 1. *Mol Biochem Parasitol* 63, 299-303.
- Laffont, C.M., Alvinerie, M., Bousquet-Melou, A., Toutain, P.L., 2001. Licking behaviour and environmental contamination arising from pour-on ivermectin for cattle. *Int J Parasitol* 31, 1687-1692.
- Laing, R., Hunt, M., Protasio, A.V., Saunders, G., Mungall, K., Laing, S., Jackson, F., Quail, M., Beech, R., Berriman, M., Gilleard, J.S., 2011. Annotation of two large contiguous regions from the *Haemonchus contortus* genome using RNA-seq and comparative analysis with *Caenorhabditis elegans*. *PLoS One* 6, e23216.
- Laing, R., Kikuchi, T., Martinelli, A., Tsai, I.J., Beech, R.N., Redman, E., Holroyd, N., Bartley, D.J., Beasley, H., Britton, C., Curran, D., Devaney, E., Gilabert, A., Hunt, M., Jackson, F., Johnston, S.L., Kryukov, I., Li, K., Morrison, A.A., Reid, A.J., Sargison, N., Saunders, G.I., Wasmuth, J.D., Wolstenholme, A., Berriman, M., Gilleard, J.S., Cotton, J.A., 2013. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome Biol* 14, R88.
- Laing, R., Maitland, K., Lecova, L., Skuce, P.J., Tait, A., Devaney, E., 2016. Analysis of putative resistance gene loci in UK field populations of *Haemonchus contortus* after 6 years of macrocyclic lactone use. *Int J Parasitol* 46, 621-630.
- Laing, S., 2010. *Caenorhabditis elegans* as a model for nematode metabolism of the anthelmintic drugs ivermectin and albendazole. PhD. University of Glasgow, Glasgow.
- Laing, S.T., Ivens, A., Laing, R., Ravikumar, S., Butler, V., Woods, D.J., Gilleard, J.S., 2010. Characterization of the xenobiotic response of *Caenorhabditis elegans* to the anthelmintic drug albendazole and the identification of novel drug glucoside metabolites. *Biochem J* 432, 505-514.
- Laing, S.T., Ivens, A., Butler, V., Ravikumar, S.P., Laing, R., Woods, D.J., Gilleard, J.S., 2012. The Transcriptional Response of *Caenorhabditis elegans* to



- Ivermectin Exposure Identifies Novel Genes Involved in the Response to Reduced Food Intake. *PLoS One* 7, e31367.
- Lamb, J., Elliott, T., Chambers, M., Chick, B., 2017. Broad spectrum anthelmintic resistance of *Haemonchus contortus* in Northern NSW of Australia. *Vet Parasitol* 241, 48-51.
- Laws, D., Genever, L. 2015. Growing and finishing lambs for better returns (Agricultural and Horticultural development board (AHDB) Beef and Lamb).
- Lawton, D.E., Reynolds, G.W., Hodgkinson, S.M., Pomroy, W.E., Simpson, H.V., 1996. Infection of sheep with adult and larval *Ostertagia circumcincta*: effects on abomasal pH and serum gastrin and pepsinogen. *Int J Parasitol* 26, 1063-1074.
- Le Jambre, L.F., 1976. Egg hatch as an *in vitro* assay of thiabendazole resistance in nematodes. *Vet Parasitol* 2, 385-391.
- Le Jambre, L.F., Southcott, W.H., Dash, K.M., 1976. Resistance of selected lines of *Haemonchus contortus* to thiabendazole, morantel tartrate and levamisole. *Int J Parasitol* 6, 217-222.
- Le Jambre, L.F., Gill, J.H., Lenane, I.J., Baker, P., 2000. Inheritance of avermectin resistance in *Haemonchus contortus*. *Int J Parasitol* 30, 105-111.
- Leal, M.L., Pivoto, F.L., Fausto, G.C., Aires, A.R., Grandó, T.H., Roos, D.H., Sudati, J.H., Wagner, C., Costa, M.M., Molento, M.B., da Rocha, J.B., 2014. Copper and selenium: auxiliary measure to control infection by *Haemonchus contortus* in lambs. *Exp Parasitol* 144, 39-43.
- Learmount, J., Stephens, N., Boughtflower, V., Barrecheuren, A., Rickell, K., 2016a. The development of anthelmintic resistance with best practice control of nematodes on commercial sheep farms in the UK. *Vet Parasitol* 229, 9-14.
- Learmount, J., Stephens, N., Boughtflower, V., Barrecheuren, A., Rickell, K., Massei, G., Taylor, M., 2016b. Three-year evaluation of best practice guidelines for nematode control on commercial sheep farms in the UK. *Vet Parasitol* 226, 116-123.
- Learmount, J., Callaby, R., Taylor, M., 2018. An observational study of ewe treatments at lambing on early infection in lambs on UK sheep farms. *Vet Parasitol* 253, 55-59.
- Leathwick, D.M., 2013. Managing anthelmintic resistance -parasite fitness, drug use strategy and the potential for reversion towards susceptibility. *Vet Parasitol* 198, 145-153.
- Leathwick, D.M., Ganesh, S., Waghorn, T.S., 2015. Evidence for reversion towards anthelmintic susceptibility in *Teladorsagia circumcincta* in response to resistance management programmes. *Int J Parasitol Drugs Drug Resist* 5, 9-15.
- Lehne, G., 2000. P-glycoprotein as a drug target in the treatment of multidrug resistant cancer. *Curr Drug Targets* 1, 85-99.
- Lespine, A., Menez, C., Bourguinat, C., Prichard, R.K., 2012. P-glycoproteins and other multidrug resistance transporters in the pharmacology of anthelmintics: Prospects for reversing transport-dependent anthelmintic resistance. *Int J Parasitol Drugs Drug Resist* 2, 58-75.
- Levin, B.R., Perrot, V., Walker, N., 2000. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genet* 154, 985-997.
- Lewis, J.A., Wu, C.H., Levine, J.H., Berg, H., 1980. Levamisole-resistant mutants of the nematode *Caenorhabditis elegans* appear to lack pharmacological acetylcholine receptors. *Neuroscience* 5, 967-989.

- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN]
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094-3100.
- Lichtenfels, J.R., Pilitt, P.A., Lancaster, M.B., 1988. Systematics of the nematodes that cause ostertagiasis in cattle, sheep and goats in North America. *Vet Parasitol* 27, 3-12.
- Lichtenfels, J.R., Hoberg, E.P., 1993. The systematics of nematodes that cause ostertagiasis in domestic and wild ruminants in North America: an update and a key to species. *Vet Parasitol* 46, 33-53.
- Little, P.R., Hodge, A., Maeder, S.J., Wirtherle, N.C., Nicholas, D.R., Cox, G.G., Conder, G.A., 2011. Efficacy of a combined oral formulation of derquantel-abamectin against the adult and larval stages of nematodes in sheep, including anthelmintic-resistant strains. *Vet Parasitol* 181, 180-193.
- Liu, Y., LeBoeuf, B., Garcia, L.R., 2007. Gαq-Coupled Muscarinic Acetylcholine Receptors Enhance Nicotinic Acetylcholine Receptor Signaling in *Caenorhabditis elegans* Mating Behavior. *J Neurosci* 27, 1411-1421.
- Lloberas, M., Alvarez, L., Entrocasso, C., Virkel, G., Ballent, M., Mate, L., Lanusse, C., Lifschitz, A., 2013. Comparative tissue pharmacokinetics and efficacy of moxidectin, abamectin and ivermectin in lambs infected with resistant nematodes: Impact of drug treatments on parasite P-glycoprotein expression. *Int J Parasitol Drugs Drug Resist* 3, 20-27.
- Lomsadze, A., Burns, P.D., Borodovsky, M., 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* 42, e119.
- Louis, E.J., Dempster, E.R., 1987. An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* 43, 805-811.
- Lovatt, F., Stubbings, L. 2018. Worm control in sheep for Better Returns (Agriculture and Horticulture Development Board (AHDB) Beef and Lamb).
- Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F., Storfer, A., 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour* 17, 142-152.
- Lubega, G.W., Klein, R.D., Geary, T.G., Prichard, R.K., 1994. *Haemonchus contortus*: The role of two  $\beta$ -tubulin gene subfamilies in the resistance to benzimidazole anthelmintics. *Biochem Pharmacol* 47, 1705-1715.
- Luo, X., Shi, X., Yuan, C., Ai, M., Ge, C., Hu, M., Feng, X., Yang, X., 2017. Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in *Haemonchus contortus*. *Parasit Vectors* 10, 31.
- Mainigi, N., Scott, M.E., Prichard, R.K., 1990. Effect of selection pressure for thiabendazole resistance on fitness of *Haemonchus contortus* in sheep. *Parasitol* 100 Pt 2, 327-335.
- Mani, T., Bourguinat, C., Keller, K., Ashraf, S., Blagburn, B., Prichard, R.K., 2016. Interaction of macrocyclic lactones with a *Dirofilaria immitis* P-glycoprotein. *Int J Parasitol* 46, 631-640.

- Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., Chen, W.-M., 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867-2873.
- Marley, C.L., Cook, R., Barrett, J., Keatinge, R., Lampkin, N.H., McBride, S.D., 2003. The effect of dietary forage on the development and survival of helminth parasites in ovine faeces. *Vet Parasitol* 118, 93-107.
- Maroso, F., Hillen, J.E.J., Pardo, B.G., Gkagkavouzis, K., Coscia, I., Hermida, M., Franch, R., Hellemans, B., Van Houdt, J., Simionati, B., Taggart, J.B., Nielsen, E.E., Maes, G., Ciavaglia, S.A., Webster, L.M.I., Volckaert, F.A.M., Martinez, P., Bargelloni, L., Ogden, R., 2018. Performance and precision of double digestion RAD (ddRAD) genotyping in large multiplexed datasets of marine fish species. *Mar Genom* 39, 64-72.
- Martin, P.J., Le Jambre, L.F., Claxton, J.H., 1981. The impact of refugia on the development of thiabendazole resistance in *Haemonchus contortus*. *Int J Parasitol* 11, 35-41.
- Martin, P.J., Anderson, N., Jarrett, R.G., 1989. Detecting benzimidazole resistance with faecal egg count reduction tests and *in vitro* assays. *Aust Vet J* 66, 236-240.
- Martin, R.J., Robertson, A.P., Buxton, S.K., Beech, R.N., Charvet, C.L., Neveu, C., 2012. Levamisole receptors: a second awakening. *Trends Parasitol* 28, 289-296.
- Martinez-Valladares, M., Donnan, A., Geldhof, P., Jackson, F., Rojo-Vazquez, F.A., Skuce, P., 2012a. Pyrosequencing analysis of the beta-tubulin gene in Spanish *Teladorsagia circumcincta* field isolates. *Vet Parasitol* 184, 371-376.
- Martinez-Valladares, M., Geldhof, P., Jonsson, N., Rojo-Vazquez, F.A., Skuce, P., 2012b. *Teladorsagia circumcincta*: Molecular characterisation of the *avr-14B* subunit and its relatively minor role in ivermectin resistance. *Int J Parasitol Drugs Drug Resist* 2, 154-161.
- Maruki, T., Lynch, M., 2017. Genotype Calling from Population-Genomic Sequencing Data. G3: Genes, Genomes, Genet 7, 1393-1404.
- Matschiner, M., Salzburger, W., 2009. TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* 25, 1982-1983.
- Mavrot, F., Hertzberg, H., Torgerson, P., 2015. Effect of gastro-intestinal nematode infection on sheep performance: a systematic review and meta-analysis. *Parasit Vectors* 8, 557-557.
- McBean, D., Nath, M., Lambe, N., Morgan-Davies, C., Kenyon, F., 2016. Viability of the Happy Factor targeted selective treatment approach on several sheep farms in Scotland. *Vet Parasitol* 218, 22-30.
- McCavera, S., Walsh, T.K., Wolstenholme, A.J., 2007. Nematode ligand-gated chloride channels: an appraisal of their involvement in macrocyclic lactone resistance and prospects for developing molecular markers. *Parasitol* 134, 1111-1121.
- McDonald, J.H., 2014. Handbook of Biological Statistics, 3rd Edition. Sparky House Publishing, Baltimore, Maryland, 94-100 pp.
- McKellar, Q.A., 1993. Interactions of *Ostertagia* species with their bovine and ovine hosts. *Int J Parasitol* 23, 451-462.
- McKenna, P.B., 1997. Further potential limitations of the undifferentiated faecal egg count reduction test for the detection of anthelmintic resistance in sheep. *N Z Vet J* 45, 244-246.

- McKenna, P.B., 2006. Further comparison of faecal egg count reduction test procedures: sensitivity and specificity. *N Z Vet J* 54, 365-366.
- McKinney, G.J., Larson, W.A., Seeb, L.W., Seeb, J.E., 2017. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). *Mol Ecol Resour* 17, 356-361.
- McMahon, C., Barley, J.P., Edgar, H.W., Ellison, S.E., Hanna, R.E., Malone, F.E., Brennan, G.P., Fairweather, I., 2013a. Anthelmintic resistance in Northern Ireland. II: Variations in nematode control practices between lowland and upland sheep flocks. *Vet Parasitol* 192, 173-182.
- McMahon, C., Bartley, D.J., Edgar, H.W., Ellison, S.E., Barley, J.P., Malone, F.E., Hanna, R.E., Brennan, G.P., Fairweather, I., 2013b. Anthelmintic resistance in Northern Ireland (I): prevalence of resistance in ovine gastrointestinal nematodes, as determined through faecal egg count reduction testing. *Vet Parasitol* 195, 122-130.
- McMahon, C., McCoy, M., Ellison, S.E., Barley, J.P., Edgar, H.W., Hanna, R.E., Malone, F.E., Brennan, G.P., Fairweather, I., 2013c. Anthelmintic resistance in Northern Ireland (III): uptake of 'SCOPS' (Sustainable Control of Parasites in Sheep) recommendations by sheep farmers. *Vet Parasitol* 193, 179-184.
- McMahon, C., Edgar, H.W.J., Barley, J.P., Hanna, R.E.B., Brennan, G.P., Fairweather, I., 2017. Control of *Nematodirus spp.* infection by sheep flock owners in Northern Ireland. *Ir Vet J* 70, 31-31.
- McQuillan, J.A., Parkhill, J., Berriman, M., Harris, S.R., Carver, T., 2011. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464-469.
- Melotti, A., Mas, C., Kuciak, M., Lorente-Trigos, A., Borges, I., Ruiz i Altaba, A., 2014. The river blindness drug Ivermectin and related macrocyclic lactones inhibit WNT-TCF pathway responses in human cancer. *EMBO Mol Med* 6, 1263-1278.
- Melville, L.A., McBean, D., Fyfe, A., Campbell, S.J., Palarea-Albaladejo, J., Kenyon, F., 2016. Effect of anthelmintic treatment strategy on strongylid nematode species composition in grazing lambs in Scotland. *Parasit Vectors* 9, 199.
- Ménez, C., Alberich, M., Kansoh, D., Blanchard, A., Lespine, A., 2016. Acquired Tolerance to Ivermectin and Moxidectin after Drug Selection Pressure in the Nematode *Caenorhabditis elegans*. *Antimicrob Agents Ch* 60, 4809-4819.
- Ménez, C., Alberich, M., Courtot, E., Guegnard, F., Blanchard, A., Aguilaniu, H., Lespine, A., 2019. The transcription factor NHR-8: A new target to increase ivermectin efficacy in nematodes. *PLoS Pathog* 15, e1007598.
- Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17, 240-248.
- Mitchell, E.S., Hunt, K.R., Wood, R., McLean, B., 2010. Anthelmintic resistance on sheep farms in Wales. *Vet Rec* 166, 650-652.
- Mitchell, S., 2016. Parasitic gastroenteritis in sheep. *Vet Rec* 179, 223-224.
- Morgan, D.O., Parnell, I.W., Rayski, C., 1951. The Seasonal Variations in the Worm Burden of Scottish Hill Sheep. *J Helminthol* 25, 177-212.
- Morgan, E.R., Hosking, B.C., Burston, S., Carder, K.M., Hyslop, A.C., Pritchard, L.J., Whitmarsh, A.K., Coles, G.C., 2012. A survey of helminth control practices on sheep farms in Great Britain and Ireland. *Vet J* 192, 390-397.

- Morgan, E.R., Charlier, J., Hendrickx, G., Biggeri, A., Catalan, D., Von Samson-Himmelstjerna, G., Demeler, J., Müller, E., Van Dijk, J., Kenyon, F., Skuce, P., Höglund, J., Kiely, P., Van Ranst, B., De Waal, T., Rinaldi, L., Cringoli, G., Hertzberg, H., Torgerson, P., Wolstenholme, A., Vercruysse, J., 2013. Global Change and Helminth Infections in Grazing Ruminants in Europe: Impacts, Trends and Sustainable Solutions. *Agriculture* 3, 484-502.
- Morris, C.A., Bisset, S.A., Vlassoff, A., Wheeler, M., West, C.J., Devantier, B.P., Mackay, A.D., 2010. Selecting for resilience in Romney sheep under nematode parasite challenge, 1994–2007. *New Zeal J Agr Res* 53, 245-261.
- Morrison, A.A., Mitchell, S., Mearns, R., Richards, I., Matthews, J.B., Bartley, D.J., 2014. Phenotypic and genotypic analysis of benzimidazole resistance in the ovine parasite *Nematodirus battus*. *Vet Res* 45, 116.
- Neveu, C., Charvet, C.L., Fauvin, A., Cortet, J., Beech, R.N., Cabaret, J., 2010. Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance. *Pharmacogenet Genomics* 20, 414-425.
- Ng'habi, K., Viana, M., Matthiopoulos, J., Lyimo, I., Killeen, G., Ferguson, H.M., 2018. Mesocosm experiments reveal the impact of mosquito control measures on malaria vector life history and population dynamics. *Sci Rep* 8, 13949-13949.
- Niciura, S.C.M., Tizioto, P.C., Moraes, C.V., Cruvinel, G.G., de Albuquerque, A.C.A., Santana, R.C.M., Chagas, A.C.d.S., Esteves, S.N., Benavides, M.V., do Amarante, A.F.T., 2019. Extreme-QTL mapping of monepantel resistance in *Haemonchus contortus*. *Parasit Vectors* 12, 403.
- Nielsen, M.K., Reinemeyer, C.R., Donecker, J.M., Leathwick, D.M., Marchiondo, A.A., Kaplan, R.M., 2014. Anthelmintic resistance in equine parasites - current evidence and knowledge gaps. *Vet Parasitol* 204, 55-63.
- Nisbet, A.J., McNeilly, T.N., Greer, A.W., Bartley, Y., Oliver, E.M., Smith, S., Palarea-Albaladejo, J., Matthews, J.B., 2016a. Protection of ewes against *Teladorsagia circumcincta* infection in the periparturient period by vaccination with recombinant antigens. *Vet Parasitol* 228, 130-136.
- Nisbet, A.J., Meeusen, E.N., González, J.F., Piedrafita, D.M., 2016b. Chapter Eight - Immunity to *Haemonchus contortus* and Vaccine Development, In: Gasser, R.B., Samson-Himmelstjerna, G.V. (Eds.) *Adv Parasit. Academic Press*, pp. 353-396.
- Njue, A.I., Hayashi, J., Kinne, L., Feng, X.P., Prichard, R.K., 2004. Mutations in the extracellular domains of glutamate-gated chloride channel alpha3 and beta subunits from ivermectin-resistant *Cooperia oncophora* affect agonist sensitivity. *J Neurochem* 89, 1137-1147.
- Njue, A.I., Prichard, R.K., 2004. Genetic variability of glutamate-gated chloride channel genes in ivermectin-susceptible and -resistant strains of *Cooperia oncophora*. *Parasitol* 129, 741-751.
- O'Connor, L.J., Walkden-Brown, S.W., Kahn, L.P., 2006. Ecology of the free-living stages of major trichostrongylid parasites of sheep. *Vet Parasitol* 142, 1-15.
- O'Connor, L.J., Walkden-Brown, S.W., Kahn, L.P., 2006. Ecology of the free-living stages of major trichostrongylid parasites of sheep. *Vet Parasitol* 142.
- Page, A.P., 2018. The sensory amphidial structures of *Caenorhabditis elegans* are involved in macrocyclic lactone uptake and anthelmintic resistance. *Int J Parasitol* 48, 1035-1042.

- Pan, X., Urban, A.E., Palejev, D., Schulz, V., Grubert, F., Hu, Y., Snyder, M., Weissman, S.M., 2008. A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *P Natl Acad Sci* 105, 15499-15504.
- Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. *Methods Ecol Evol* 8, 1360-1373.
- Park, S.T., Kim, J., 2016. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *Int Neurourol J* 20, S76-83.
- Park, Y.S., Kim, S., Shin, Y., Choi, B., Cho, N.J., 2003. Alternative splicing of the muscarinic acetylcholine receptor GAR-3 in *Caenorhabditis elegans*. *Biochem Biophys Res Commun* 308, 961-965.
- Patel, M.R., 1997. Effects of Ivermectin on Eggs and First-Stage Larvae of Nematodes. *Bios* 68, 152-162.
- Peden, A.S., Mac, P., Fei, Y.-J., Castro, C., Jiang, G., Murfitt, K.J., Miska, E.A., Griffin, J.L., Ganapathy, V., Jorgensen, E.M., 2013. Betaine acts on a ligand-gated ion channel in the nervous system of the nematode *C. elegans*. *Nat Neurosci* 16, 1794.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7, e37135.
- Phair, N.L., Toonen, R.J., Knapp, I., von der Heyden, S., 2019. Shared genomic outliers across two divergent population clusters of a highly threatened seagrass. *PeerJ* 7, e6806.
- Phillips, M.A., Long, A.D., Greenspan, Z.S., Greer, L.F., Burke, M.K., Villeponteau, B., Matsagas, K.C., Rizza, C.L., Mueller, L.D., Rose, M.R., 2016. Genome-wide analysis of long-term evolutionary domestication in *Drosophila melanogaster*. *Sci Rep* 6, 39281.
- Pica-Mattoccia, L., Carlini, D., Guidi, A., Cimica, V., Vigorosi, F., Cioli, D., 2006. The schistosome enzyme that activates oxamniquine has the characteristics of a sulfotransferase. *Mem Inst Oswaldo Cruz* 101 Suppl 1, 307-312.
- Prichard, R.K., Roulet, A., 2007. ABC transporters and beta-tubulin in macrocyclic lactone resistance: prospects for marker development. *Parasitol* 134, 1123-1132.
- Pryszcz, L.P., Gabaldón, T., 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* 44, e113-e113.
- Puttachary, S., Trailovic, S.M., Robertson, A.P., Thompson, D.P., Woods, D.J., Martin, R.J., 2013. Derquantel and abamectin: effects and interactions on isolated tissues of *Ascaris suum*. *Mol Biochem Parasitol* 188, 79-86.
- Qian, H., Martin, R.J., Robertson, A.P., 2006. Pharmacology of N-, L-, and B-subtypes of nematode nAChR resolved at the single-channel level in *Ascaris suum*. *FASEB J* 20, 2606-2608.
- Qian, H., Robertson, A.P., Powell-Coffman, J.A., Martin, R.J., 2008. Levamisole resistance resolved at the single-channel level in *Caenorhabditis elegans*. *FASEB J* 22, 3247-3254.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.
- Quinlan, A.R., 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47, 11.12.11-34.
- Ramünke, S., Melville, L., Rinaldi, L., Hertzberg, H., de Waal, T., von Samson-Himmelstjerna, G., Cringoli, G., Mavrot, F., Skuce, P., Krücken, J., Demeler, J., 2016. Benzimidazole resistance survey for *Haemonchus*, *Teladorsagia* and *Trichostrongylus* in three European countries using

- pyrosequencing including the development of new assays for *Trichostrongylus*. *Int J Parasitol Drugs Drug Resist* 6, 230-240.
- Ranjan, S., Wang, G.T., Hirschlein, C., Simkins, K.L., 2002. Selection for resistance to macrocyclic lactones by *Haemonchus contortus* in sheep. *Vet Parasitol* 103, 109-117.
- Ravindranath, A.K., Kaur, S., Wernyj, R.P., Kumaran, M.N., Miletti-Gonzalez, K.E., Chan, R., Lim, E., Madura, K., Rodriguez-Rodriguez, L., 2015. CD44 promotes multi-drug resistance by protecting P-glycoprotein from FBXO21-mediated ubiquitination. *Oncotarget* 6, 26308-26321.
- Raza, A., Bagnall, N.H., Jabbar, A., Kopp, S.R., Kotze, A.C., 2016a. Increased expression of ATP binding cassette transporter genes following exposure of *Haemonchus contortus* larvae to a high concentration of monepantel *in vitro*. *Parasit Vectors* 9, 522.
- Raza, A., Kopp, S.R., Bagnall, N.H., Jabbar, A., Kotze, A.C., 2016b. Effects of *in vitro* exposure to ivermectin and levamisole on the expression patterns of ABC transporters in *Haemonchus contortus* larvae. *Int J Parasitol Drugs Drug Resist* 6, 103-115.
- Raza, A., Lamb, J., Chambers, M., Hunt, P.W., Kotze, A.C., 2016c. Larval development assays reveal the presence of sub-populations showing high- and low-level resistance in a monepantel (Zolvix®)-resistant isolate of *Haemonchus contortus*. *Vet Parasitol* 220, 77-82.
- Redman, E., Packard, E., Grillo, V., Smith, J., Jackson, F., Gilleard, J.S., 2008. Microsatellite analysis reveals marked genetic differentiation between *Haemonchus contortus* laboratory isolates and provides a rapid system of genetic fingerprinting. *Int J Parasitol* 38, 111-122.
- Redman, E., Sargison, N., Whitelaw, F., Jackson, F., Morrison, A., Bartley, D.J., Gilleard, J.S., 2012. Introgression of ivermectin resistance genes into a susceptible *Haemonchus contortus* strain by multiple backcrossing. *PLoS Pathog* 8, e1002534.
- Redman, E., Whitelaw, F., Tait, A., Burgess, C., Bartley, Y., Skuce, P.J., Jackson, F., Gilleard, J.S., 2015. The emergence of resistance to the benzimidazole anthelmintics in parasitic nematodes of livestock is characterised by multiple independent hard and soft selective sweeps. *PLoS Negl Trop Dis* 9, e0003494.
- Rees, E., Phillips, K. 2016. BRP+ Breeding Ewe Lambs (Agriculture and Horticulture Development Board (AHDB) Beef and Lamb).
- Reid, J.F., Armour, J., 1975. Seasonal variations in the gastro-intestinal nematode populations of Scottish hill sheep. *Res Vet Sci* 18, 307-313.
- Reinecke, R.K., Bruckner, C.M., De Villiers, I.L., 1980. Studies on *Haemonchus contortus*. III. Titration of *Trichostrongylus axei* and expulsion of *H. contortus*. *Onderstepoort J Vet Res* 47, 35-44.
- Rezansoff, A.M., Laing, R., Gilleard, J.S., 2016. Evidence from two independent backcross experiments supports genetic linkage of microsatellite Hcms8a20, but not other candidate loci, to a major ivermectin resistance locus in *Haemonchus contortus*. *Int J Parasitol* 46, 653-661.
- Rinaldi, L., Coles, G.C., Maurelli, M.P., Musella, V., Cringoli, G., 2011. Calibration and diagnostic accuracy of simple flotation, McMaster and FLOTAC for parasite egg counts in sheep. *Vet Parasitol* 177, 345-352.
- Rinaldi, L., Levecke, B., Bosco, A., Ianniello, D., Pepe, P., Charlier, J., Cringoli, G., Vercruysse, J., 2014. Comparison of individual and pooled faecal samples in sheep for the assessment of gastrointestinal strongyle infection intensity

- and anthelmintic drug efficacy using McMaster and Mini-FLOTAC. *Vet Parasitol* 205, 216-223.
- Roach, M.J., Schmidt, S.A., Borneman, A.R., 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19, 460.
- Robasky, K., Lewis, N.E., Church, G.M., 2014. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* 15, 56-62.
- Robinson, M.W., McFerran, N., Trudgett, A., Hoey, L., Fairweather, I., 2004. A possible model of benzimidazole binding to beta-tubulin disclosed by invoking an inter-domain movement. *J Mol Graph Model* 23, 275-284.
- Rochette, N.C., Catchen, J.M., 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat Protoc* 12, 2640-2659.
- Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol* 0.
- Rockman, M.V., Kruglyak, L., 2009. Recombinational Landscape and Population Genomics of *Caenorhabditis elegans*. *PLoS Genet* 5, e1000419.
- Rode, N.O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J., Gay, L., 2018. How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Mol Ecol Resour* 18, 194-203.
- Roeber, F., Jex, A.R., Campbell, A.J., Nielsen, R., Anderson, G.A., Stanley, K.K., Gasser, R.B., 2012. Establishment of a robotic, high-throughput platform for the specific diagnosis of gastrointestinal nematode infections in sheep. *Int J Parasitol* 42, 1151-1158.
- Roeber, F., Jex, A.R., Gasser, R.B., 2013a. Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for exploring epidemiology and drug resistance - an Australian perspective. *Parasit Vectors* 6, 153.
- Roeber, F., Jex, A.R., Gasser, R.B., 2013b. Advances in the diagnosis of key gastrointestinal nematode infections of livestock, with an emphasis on small ruminants. *Biotechnol Adv* 31.
- Roeber, F., Morrison, A., Casaert, S., Smith, L., Claerebout, E., Skuce, P., 2017. Multiplexed-tandem PCR for the specific diagnosis of gastrointestinal nematode infections in sheep: an European validation study. *Parasit Vectors* 10, 226.
- Rose, H., Rinaldi, L., Bosco, A., Mavrot, F., de Waal, T., Skuce, P., Charlier, J., Torgerson, P.R., Hertzberg, H., Hendrickx, G., Vercruyse, J., Morgan, E.R., 2015. Widespread anthelmintic resistance in European farmed ruminants: a systematic review. *Vet Rec* 176, 546.
- Rose, H., Caminade, C., Bolajoko, M.B., Phelan, P., Dijk, J.v., Baylis, M., Williams, D., Morgan, E.R., van Dijk, J., 2016. Climate-driven changes to the spatio-temporal distribution of the parasitic nematode, *Haemonchus contortus*, in sheep in Europe. *Global Change Biol* 22, 1271-1285.
- Ross, J.G., 1970. Genetic differences in the susceptibility of sheep to infection with *Trichostrongylus axei*. A comparison of Scottish Black-face and Dorset breeds. *Res Vet Sci* 11, 465-468.
- Rostant, W.G., Wedell, N., Hosken, D.J., 2012. Transposable elements and insecticide resistance. *Adv Genet* 78, 169-201.
- Roy, E.A., Hoste, H., Beveridge, I., 2004. The effects of concurrent experimental infections of sheep with *Trichostrongylus colubriformis* and *T. vitrinus* on nematode distributions, numbers and on pathological changes. *Parasite* 11, 293-300.



- RStudio-Team, 2015. RStudio: Integrated Development for R. RStudio, Inc, Boston, MA.
- Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring Phylogenies from RAD Sequence Data. *PLoS One* 7, e33394.
- Rufener, L., Kaminsky, R., Mäser, P., 2009a. *In vitro* selection of *Haemonchus contortus* for benzimidazole resistance reveals a mutation at amino acid 198 of  $\beta$ -tubulin. *Mol Biochem Parasitol* 168, 120-122.
- Rufener, L., Mäser, P., Roditi, I., Kaminsky, R., 2009b. *Haemonchus contortus* Acetylcholine Receptors of the DEG-3 Subfamily and Their Role in Sensitivity to Monepantel. *PLoS Pathog* 5, e1000380.
- Saccareau, M., Salle, G., Robert-Granie, C., Duchemin, T., Jacquiet, P., Blanchard, A., Cabaret, J., Moreno, C.R., 2017. Meta-analysis of the parasitic phase traits of *Haemonchus contortus* infection in sheep. *Parasit Vectors* 10, 201.
- Sakkas, P., Houdijk, J.G., Athanasiadou, S., Kyriazakis, I., 2012. Sensitivity of periparturient breakdown of immunity to parasites to dietary protein source. *J Anim Sci* 90, 3954-3962.
- Salih, N.E., Grainger, J.N.R., 1982. The effect of constant and changing temperatures on the development of the eggs and larvae of *Ostertagia circumcincta*. *J Therm Biol* 7, 35-38.
- Salisbury, J.R., Arundel, J.H., 1970. Peri-parturient deposition of nematode eggs by ewes and residual pasture contamination as sources of infection for lambs. *Aust Vet J* 46.
- Sallé, G., Laing, R., Cotton, J.A., Maitland, K., Martinelli, A., Holroyd, N., Tracey, A., Berriman, M., Smith, W.D., Newlands, G.F.J., Hanks, E., Devaney, E., Britton, C., 2018. Transcriptomic profiling of nematode parasites surviving vaccine exposure. *Int J Parasitol* 48, 395-402.
- Sallé, G., Doyle, S.R., Cortet, J., Cabaret, J., Berriman, M., Holroyd, N., Cotton, J.A., 2019. The global diversity of *Haemonchus contortus* is shaped by human intervention and climate. *Nat Commun* 10, 4811.
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463-5467.
- Sangster, N.C., Whitlock, H.V., Russ, I.G., Gunawan, M., Griffin, D.L., Kelly, J.D., 1979. *Trichostrongylus colubriformis* and *Ostertagia circumcincta* resistant to levamisole, morantel tartrate and thiabendazole: occurrence of field strains. *Res Vet Sci* 27, 106-110.
- Sangster, N.C., Riley, F.L., Collins, G.H., 1988. Investigation of the mechanism of levamisole resistance trichostrongylid nematodes of sheep. *Int J Parasitol* 18, 813-818.
- Sangster, N.C., Bjorn, H., 1995. Levamisole resistance in *Haemonchus contortus* selected at different stages of infection. *Int J Parasitol* 25, 343-348.
- Sangster, N.C., Redwin, J.M., Bjorn, H., 1998. Inheritance of levamisole and benzimidazole resistance in an isolate of *Haemonchus contortus*. *Int J Parasitol* 28, 503-510.
- Sangster, N.C., 1999. Anthelmintic resistance: past, present and future. *Int J Parasitol* 29, 115-124.
- Sarai, R.S., Kopp, S.R., Coleman, G.T., Kotze, A.C., 2013. Acetylcholine receptor subunit and P-glycoprotein transcription patterns in levamisole-susceptible and -resistant *Haemonchus contortus*. *Int J Parasitol Drugs Drug Resist* 3, 51-58.
- Sarai, R.S., Kopp, S.R., Coleman, G.T., Kotze, A.C., 2014. Drug-efflux and target-site gene expression patterns in *Haemonchus contortus* larvae able to

- survive increasing concentrations of levamisole *in vitro*. Int J Parasitol Drugs Drug Resist 4, 77-84.
- Sargison, N., Scott, P., Jackson, F., 2001. Multiple anthelmintic resistance in sheep. Vet Rec 149, 778-779.
- Sargison, N., 2011. Responsible use of anthelmintics for nematode control in sheep and cattle. In Practice 33, 318-327.
- Sargison, N., Redman, E., Morrison, A., Bartley, D.J., Jackson, F., Naghra, H., Holroyd, N.E., Berriman, M., Cotton, J., Gilleard, J.S., 2017. A method for single pair mating in an obligate parasitic nematode. Int J Parasitol 48, 159-165.
- Sargison, N.D., Scott, P.R., Jackson, F., 2002. Teladorsagiosis in young lambs and extended postparturient susceptibility in moxidectin-treated ewes grazing heavily contaminated pastures. Vet Rec 151, 353-355.
- Sargison, N.D., Jackson, F., Bartley, D.J., Moir, A.C., 2005. Failure of moxidectin to control benzimidazole-, levamisole- and ivermectin-resistant *Teladorsagia circumcincta* in a sheep flock. Vet Rec 156, 105-109.
- Sargison, N.D., Jackson, F., Bartley, D.J., Wilson, D.J., Stenhouse, L.J., Penny, C.D., 2007. Observations on the emergence of multiple anthelmintic resistance in sheep flocks in the south-east of Scotland. Vet Parasitol 145, 65-76.
- Sargison, N.D., 2009. Development of genetic crossing methods to identify genes associated with macrocyclic lactone resistance in the sheep nematode parasite, *Haemonchus contortus*. PhD. Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh.
- Sargison, N.D., Jackson, F., Wilson, D.J., Bartley, D.J., Penny, C.D., Gilleard, J.S., 2010. Characterisation of milbemycin-, avermectin-, imidazothiazole- and benzimidazole-resistant *Teladorsagia circumcincta* from a sheep flock. Vet Rec 166, 681-686.
- Sargison, N.D., Bartram, D.J., Wilson, D.J., 2012. Use of a long acting injectable formulation of moxidectin to control the periparturient rise in faecal *Teladorsagia circumcincta* egg output of ewes. Vet Parasitol 189, 274-283.
- Sargison, N.D., 2014. Sustainable helminth control practices in the United Kingdom. Small Ruminant Res 118, 35-40.
- Schlenke, T.A., Begun, D.J., 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proc Natl Acad Sci U S A 101, 1626-1631.
- Schulman, M., Doherty, P., Arison, B., 1993. Microbial conversion of avermectins by *Saccharopolyspora erythraea*: glycosylation at C-4' and C-4". Antimicrob Agents Chemother 37, 1737-1741.
- SCOPS 2019. Know your Anthelmintics Groups (<https://www.scops.org.uk/workspace/pdfs/know-your-anthelmintic-groups.pdf>).
- Scott, E.W., Baxter, P., Armour, J., 1991. Fecundity of anthelmintic resistant adult *Haemonchus contortus* after exposure to ivermectin or benzimidazoles *in vivo*. Res Vet Sci 50, 247-249.
- Scott, I., Pomroy, W.E., Kenyon, P.R., Smith, G., Adlington, B., Moss, A., 2013. Lack of efficacy of monepantel against *Teladorsagia circumcincta* and *Trichostrongylus colubriformis*. Vet Parasitol 198, 166-171.
- Sellings, L., Pereira, S., Qian, C., Dixon-McDougall, T., Nowak, C., Zhao, B., Tyndale, R.F., van der Kooy, D., 2013. Nicotine-motivated behavior in *Caenorhabditis elegans* requires the nicotinic acetylcholine receptor subunits *acr-5* and *acr-15*. Eur J Neurosci 37, 743-756.

- Sheps, J.A., Ralph, S., Zhao, Z., Baillie, D.L., Ling, V., 2004. The ABC transporter gene family of *Caenorhabditis elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. *Genome Biol* 5, R15.
- Shortt, J.A., Card, D.C., Schield, D.R., Liu, Y., Zhong, B., Castoe, T.A., Carlton, E.J., Pollock, D.D., 2017. Whole Genome Amplification and Reduced-Representation Genome Sequencing of *Schistosoma japonicum* Miracidia. *PLoS Negl Trop Dis* 11, e0005292.
- Sieburth, D., Ch'ng, Q., Dybbs, M., Tavazoie, M., Kennedy, S., Wang, D., Dupuy, D., Rual, J.-F., Hill, D.E., Vidal, M., Ruvkun, G., Kaplan, J.M., 2005. Systematic analysis of genes required for synapse structure and function. *Nat* 436, 510-517.
- Silvestre, A., Chartier, C., Sauve, C., Cabaret, J., 2000. Relationship between helminth species diversity, intensity of infection and breeding management in dairy goats. *Vet Parasitol* 94, 91-105.
- Silvestre, A., Humbert, J.F., 2000. A molecular tool for species identification and benzimidazole resistance diagnosis in larval communities of small ruminant parasites. *Exp Parasitol* 95, 271-276.
- Silvestre, A., Cabaret, J., 2002. Mutation in position 167 of isotype 1 beta-tubulin gene of Trichostrongylid nematodes: role in benzimidazole resistance? *Mol Biochem Parasitol* 120, 297-300.
- Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210-3212.
- Sinclair, R., Melville, L., Sargison, F., Kenyon, F., Nussey, D., Watt, K., Sargison, N., 2016. Gastrointestinal nematode species diversity in Soay sheep kept in a natural environment without active parasite control. *Vet Parasitol* 227, 1-7.
- Skuce, P., Stenhouse, L., Jackson, F., Hypsa, V., Gilleard, J., 2010. Benzimidazole resistance allele haplotype diversity in United Kingdom isolates of *Teladorsagia circumcincta* supports a hypothesis of multiple origins of resistance by recurrent mutation. *Int J Parasitol* 40, 1247-1255.
- Sloan, M.A., Reaves, B.J., Maclean, M.J., Storey, B.E., Wolstenholme, A.J., 2015. Expression of nicotinic acetylcholine receptor subunits from parasitic nematodes in *Caenorhabditis elegans*. *Mol Biochem Parasitol* 204, 44-50.
- Smeal, M.G., Gough, P.A., Jackson, A.R., Hotson, I.K., 1968. The occurrence of strains of *Haemonchus contortus* resistant to thiabendazole. *Aust Vet J* 44, 108-109.
- Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M.H., Baldock, R., Barbiera, G., Bardou, P., Beck, T., Blake, A., Bonierbale, M., Brookes, A.J., Bucci, G., Buetti, I., Burge, S., Cabau, C., Carlson, J.W., Chelala, C., Chrysostomou, C., Cittaro, D., Collin, O., Cordova, R., Cutts, R.J., Dassi, E., Genova, A.D., Djari, A., Esposito, A., Estrella, H., Eyraas, E., Fernandez-Banet, J., Forbes, S., Free, R.C., Fujisawa, T., Gadaleta, E., Garcia-Manteiga, J.M., Goodstein, D., Gray, K., Guerra-Assunção, J.A., Haggarty, B., Han, D.-J., Han, B.W., Harris, T., Harshbarger, J., Hastings, R.K., Hayes, R.D., Hoede, C., Hu, S., Hu, Z.-L., Hutchins, L., Kan, Z., Kawaji, H., Keliét, A., Kerhornou, A., Kim, S., Kinsella, R., Klopp, C., Kong, L., Lawson, D., Lazarevic, D., Lee, J.-H., Letellier, T., Li, C.-Y., Lio, P., Liu, C.-J., Luo, J., Maass, A., Mariette, J., Maurel, T., Merella, S., Mohamed, A.M., Moreews, F., Nabihoudine, I., Ndegwa, N., Noirot, C., Perez-Llamas, C., Primig, M., Quattrone, A., Quesneville, H., Rambaldi, D., Reecy, J., Riba, M., Rosanoff, S., Saddiq, A.A., Salas, E., Sallou, O., Shepherd, R., Simon, R., Sperling, L., Spooner, W., Staines,

- D.M., Steinbach, D., Stone, K., Stupka, E., Teague, J.W., Dayem Ullah, A.Z., Wang, J., Ware, D., Wong-Erasmus, M., Youens-Clark, K., Zadissa, A., Zhang, S.-J., Kasprzyk, A., 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 43, W589-W598.
- Smit, R.B., Schnabel, R., Gaudet, J., 2008. The HLH-6 Transcription Factor Regulates *C. elegans* Pharyngeal Gland Development and Function. *PLoS Genet* 4, e1000222.
- Smith, J.M., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genet Res* 23, 23-35.
- Smith, W.D., Jackson, F., Jackson, E., Williams, J., 1985. Age immunity to *Ostertagia circumcincta*: comparison of the local immune responses of 4 1/2- and 10-month-old lambs. *J Comp Pathol* 95, 235-245.
- Smith, W.D., Pettit, D., Smith, S.K., 2001. Cross-protection studies with gut membrane glycoprotein antigens from *Haemonchus contortus* and *Teladorsagia circumcincta*. *Parasite Immunol* 23, 203-211.
- Smith, W.D., 2007. Some observations on immunologically mediated inhibited *Teladorsagia circumcincta* and their subsequent resumption of development in sheep. *Vet Parasitol* 147, 103-109.
- Smus, J.P., Ludlow, E., Dalliere, N., Luedtke, S., Monfort, T., Lilley, C., Urwin, P., Walker, R.J., O'Connor, V., Holden-Dye, L., Mahajan, S., 2017. Coherent anti-Stokes Raman scattering (CARS) spectroscopy in *Caenorhabditis elegans* and *Globodera pallida*: evidence for an ivermectin-activated decrease in lipid stores. *Pest Manag Sci* 73, 2550-2558.
- Sommerville, R.I., 1953. Development of *Ostertagia circumcincta* in the Abomasal Mucosa of the Sheep. *Nat* 171, 482-483.
- Spanier, B., Stürzenbaum, S.R., Holden-Dye, L.M., Baumeister, R., 2005. *Caenorhabditis elegans* Neprilysin NEP-1: an Effector of Locomotion and Pharyngeal Pumping. *J Mol Biol* 352, 429-437.
- Stadalienė, I., Höglund, J., Petkevičius, S., 2015. Seasonal patterns of gastrointestinal nematode infection in goats on two Lithuanian farms. *Acta Vet Scand* 57, 16-16.
- Stanke, M., Schoffmann, O., Morgenstern, B., Waack, S., 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62.
- Stanke, M., Diekhans, M., Baertsch, R., Haussler, D., 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637-644.
- Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., Smadja, C.M., 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc Lond B Biol Sci* 372.
- Stear, M.J., Bishop, S.C., 1999. The curvilinear relationship between worm length and fecundity of *Teladorsagia circumcincta*. *Int J Parasitol* 29, 777-780.
- Stear, M.J., Strain, S., Bishop, S.C., 1999. How lambs control infection with *Ostertagia circumcincta*. *Vet Immunol Immunopathol* 72, 213-218.
- Steger, K.A., Avery, L., 2004. The GAR-3 muscarinic receptor cooperates with calcium signals to regulate muscle contraction in the *Caenorhabditis elegans* pharynx. *Genet* 167, 633-643.
- Stephens, A.S., Stephens, S.R., Morrison, N.A., 2011. Internal control genes for quantitative RT-PCR expression analysis in mouse osteoblasts, osteoclasts and macrophages. *BMC Res Notes* 4, 410-410.

- Storey, B., Marcellino, C., Miller, M., Maclean, M., Mostafa, E., Howell, S., Sakanari, J., Wolstenholme, A., Kaplan, R., 2014. Utilization of computer processed high definition video imaging for measuring motility of microscopic nematode stages on a quantitative scale: "The Worminator". *Int J Parasitol Drugs Drug Resist* 4, 233-243.
- Strube, C., Haake, C., Sager, H., Schorderet Weber, S., Kaminsky, R., Buschbaum, S., Joekel, D., Schicht, S., Kremmer, E., Korrell, J., Schnieder, T., von Samson-Himmelstjerna, G., 2015. Vaccination with recombinant paramyosin against the bovine lungworm *Dictyocaulus viviparus* considerably reduces worm burden and larvae shedding. *Parasit Vectors* 8, 119.
- Stuchlikova, L.R., Matouskova, P., Vokral, I., Lamka, J., Szotakova, B., Seckarova, A., Dimunova, D., Nguyen, L.T., Varady, M., Skalova, L., 2018. Metabolism of albendazole, ricobendazole and flubendazole in *Haemonchus contortus* adults: Sex differences, resistance-related differences and the identification of new metabolites. *Int J Parasitol Drugs Drug Resist* 8, 50-58.
- Sturgess-Osborne, C., Burgess, S., Mitchell, S., Wall, R., 2019. Multiple resistance to macrocyclic lactones in the sheep scab mite *Psoroptes ovis*. *Vet Parasitol* 272, 79-82.
- Sutherland, I.A., Leathwick, D.M., Moen, I.C., Bisset, S.A., 2002. Resistance to therapeutic treatment with macrocyclic lactone anthelmintics in *Ostertagia circumcincta*. *Vet Parasitol* 109, 91-99.
- Sutherland, I.A., Brown, A.E., Leathwick, D.M., 2003a. The Effect of Anthelmintic Capsules on the Egg Output and Larval Viability of Drug-resistant Parasites. *Vet Res Commun* 27, 149-157.
- Sutherland, I.A., Brown, A.E., Leathwick, D.M., Bisset, S.A., 2003b. Resistance to prophylactic treatment with macrocyclic lactone anthelmintics in *Teladorsagia circumcincta*. *Vet Parasitol* 115, 301-309.
- Sutherland, I.H., Campbell, W.C., 1990. Development, pharmacokinetics and mode of action of ivermectin. *Acta Leiden* 59, 161-168.
- Swerdlow, H., Gesteland, R., 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res* 18, 1415-1419.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genet* 123, 585-595.
- Takala-Harrison, S., Clark, T.G., Jacob, C.G., Cummings, M.P., Miotto, O., Dondorp, A.M., Fukuda, M.M., Nosten, F., Noedl, H., Imwong, M., Bethell, D., Se, Y., Lon, C., Tyner, S.D., Saunders, D.L., Socheat, D., Arie, F., Phyo, A.P., Starzengruber, P., Fuehrer, H.-P., Swoboda, P., Stepniewska, K., Flegg, J., Arze, C., Cerqueira, G.C., Silva, J.C., Ricklefs, S.M., Porcella, S.F., Stephens, R.M., Adams, M., Kenefic, L.J., Campino, S., Auburn, S., MacInnis, B., Kwiatkowski, D.P., Su, X.-z., White, N.J., Ringwald, P., Plowe, C.V., 2013. Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *P Natl Acad Sci* 110, 240-245.
- Tan, J.C., Miller, B.A., Tan, A., Patel, J.J., Cheeseman, I.H., Anderson, T.J., Manske, M., Maslen, G., Kwiatkowski, D.P., Ferdig, M.T., 2011. An optimized microarray platform for assaying genomic variation in *Plasmodium falciparum* field populations. *Genome Biol* 12, R35.
- Taylor, D.M., Thomas, R.J., 1986. The development of immunity to *Nematodirus battus* in lambs. *Int J Parasitol* 16.

- Taylor, M., Sargison, N., Howe, M., Scott, P., 2016. Dosing ewes at lambing time. *Vet Rec* 178, 377-378.
- Taylor, M.A., 1990. A larval development test for the detection of anthelmintic resistance in nematodes of sheep. *Res Vet Sci* 49, 198-202.
- Taylor, S.M., Pearson, G.R., 1979a. *Trichostrongylus vitrinus* in sheep: I. The location of nematodes during parasitic development and associated pathological changes in the small intestine. *J Comp Pathol* 89, 397-403.
- Taylor, S.M., Pearson, G.R., 1979b. *Trichostrongylus vitrinus* in sheep. II. The location of nematodes and associated pathological changes in the small intestine during clinical infection. *J Comp Pathol* 89, 405-412.
- Taylor, S.M., Kenny, J., Edgar, H.W., Ellison, S., Ferguson, L., 1997. Efficacy of moxidectin, ivermectin and albendazole oral drenches for suppression of periparturient rise in ewe worm egg output and reduction of anthelmintic treatment for lambs. *Vet Rec* 141, 357-360.
- Thienpont, D., Vanparijs, O.F.J., Raeymaekers, A.H.M., Vandenberg, J., Demoen, P.J.A., Allewijn, F.T.N., Marsboom, R.P.H., Niemegeers, C.J.E., Schellekens, K.H.L., Janssen, P.A.J., 1966. Tetramisole (R 8299), A New, Potent Broad Spectrum Anthelmintic. *Nat* 209, 1084-1086.
- Thomas, D.R., 1991. The epidemiology of *Nematodirus battus* - is it changing? *Parasitol* 102 Pt 1, 147-155.
- Torgerson, P.R., Paul, M., Furrer, R., 2014. Evaluating faecal egg count reduction using a specifically designed package “eggCounts” in R and a user friendly web interface. *Int J for Parasitol* 44, 299-303.
- Towers, P.R., Edwards, B., Richmond, J.E., Sattelle, D.B., 2005. The *Caenorhabditis elegans lev-8* gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit. *J Neurochem* 93, 1-9.
- Treinin, M., Chalfie, M., 1995. A mutated acetylcholine receptor subunit causes neuronal degeneration in *C. elegans*. *Neuron* 14, 871-877.
- Treinin, M., Gillo, B., Liebman, L., Chalfie, M., 1998. Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. *Proc Natl Acad Sci U S A* 95, 15492-15495.
- Trost, B., Walker, S., Wang, Z., Thiruvahindrapuram, B., MacDonald, J.R., Sung, W.W.L., Pereira, S.L., Whitney, J., Chan, A.J.S., Pellicchia, G., Reuter, M.S., Lok, S., Yuen, R.K.C., Marshall, C.R., Merico, D., Scherer, S.W., 2018. A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. *Am J Hum Genet* 102, 142-155.
- Tsai, I.J., Hunt, M., Holroyd, N., Huckvale, T., Berriman, M., Kikuchi, T., 2014. Summarizing specific profiles in Illumina sequencing from whole-genome amplified DNA. *DNA Res* 21, 243-254.
- Turnbull, F., Jonsson, N.N., Kenyon, F., Skuce, P.J., Bisset, S.A., 2018. P-glycoprotein-9 and macrocyclic lactone resistance status in selected strains of the ovine gastrointestinal nematode, *Teladorsagia circumcincta*. *Int J Parasitol Drugs Drug Resist* 8, 70-80.
- Turnbull, F., Devaney, E., Morrison, A.A., Laing, R., Bartley, D.J., 2019. Genotypic characterisation of monepantel resistance in historical and newly derived field strains of *Teladorsagia circumcincta*. *Int J Parasitol Drugs Drug Resist* 11, 59-69.
- Tyrrell, K.L., Dobson, R.J., Stein, P.A., Walkden-Brown, S.W., 2002. The effects of ivermectin and moxidectin on egg viability and larval development of ivermectin-resistant *Haemonchus contortus*. *Vet Parasitol* 107, 85-93.

- Urdaneta-Marquez, L., Bae, S.H., Janukavicius, P., Beech, R., Dent, J., Prichard, R., 2014. A *dyf-7* haplotype causes sensory neuron defects and is associated with macrocyclic lactone resistance worldwide in the nematode parasite *Haemonchus contortus*. *Int J Parasitol* 44, 1063-1071.
- Valentim, C.L.L., Cioli, D., Chevalier, F.D., Cao, X., Taylor, A.B., Holloway, S.P., Pica-Mattocchia, L., Guidi, A., Basso, A., Tsai, I.J., Berriman, M., Carvalho-Queiroz, C., Almeida, M., Aguilar, H., Frantz, D.E., Hart, P.J., LoVerde, P.T., Anderson, T.J.C., 2013. Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. *Science* 342, 1385-1389.
- Van den Brom, R., Moll, L., Kappert, C., Vellema, P., 2015. *Haemonchus contortus* resistance to monepantel in sheep. *Vet Parasitol* 209, 278-280.
- van Dijk, J., de Louw, M.D., Kalis, L.P., Morgan, E.R., 2009. Ultraviolet light increases mortality of nematode larvae and can explain patterns of larval availability at pasture. *Int J Parasitol* 39, 1151-1156.
- van Dijk, J., Morgan, E.R., 2011. The influence of water on the migration of infective trichostrongyloid larvae onto grass. *Parasitol* 138, 780-788.
- van Wyk, J.A., Malan, F.S., 1988. Resistance of field strains of *Haemonchus contortus* to ivermectin, closantel, radoxanide and the benzimidazoles in South Africa. *Vet Rec* 123, 226-228.
- van Wyk, J.A., 2001. Refugia - overlooked as perhaps the most potent factor concerning the development of anthelmintic resistance. *Onderstepoort J Vet Res* 68, 55-67.
- van Wyk, J.A., Meyer, J., van Rensburg, L.J., 2003. Little anthelmintic drenching required for lambs finished over winter and spring on annually re-established irrigated pasture on the Highveld of Gauteng Province, South Africa. *J S Afr Vet Assoc* 74, 2-6.
- van Wyk, J.A., Mayhew, E., 2013. Morphological identification of parasitic nematode infective larvae of small ruminants and cattle: A practical lab guide. 2013 80.
- Varady, M., Biorn, H., Nansen, P., 1996. *In vitro* characterization of anthelmintic susceptibility of field isolates of the pig nodular worm *Oesophagostomum spp.*, susceptible or resistant to various anthelmintics. *Int J Parasitol* 26, 733-740.
- Varady, M., Corba, J., 1999. Comparison of six *in vitro* tests in determining benzimidazole and levamisole resistance in *Haemonchus contortus* and *Ostertagia circumcincta* of sheep. *Vet Parasitol* 80, 239-249.
- Vera-Maloof, F.Z., Saavedra-Rodriguez, K., Penilla-Navarro, R.P., Rodriguez-Ramirez, A.D., Dzul, F., Manrique, P., Black, W.C., 2019. Loss of pyrethroid resistance in newly established laboratory colonies of *Aedes aegypti*. *bioRxiv*, 760710.
- VokřÁL, I., BártíKová, H., Prchal, L., StuchlíKová, L., SkÁLovÁ, L., SzotÁKová, B., Lamka, J., VÁRady, M., KubíČEK, V., 2012. The metabolism of flubendazole and the activities of selected biotransformation enzymes in *Haemonchus contortus* strains susceptible and resistant to anthelmintics. *Parasitol* 139, 1309-1316.
- von Samson-Himmelstjerna, G., Coles, G., Jackson, F., Bauer, C., Borgsteede, F., Cirak, V., Demeler, J., Donnan, A., Dorny, P., Epe, C., 2009a. Standardization of the egg hatch test for the detection of benzimidazole resistance in parasitic nematodes. *Parasitol Res* 105.
- von Samson-Himmelstjerna, G., Walsh, T.K., Donnan, A.A., Carriere, S., Jackson, F., Skuce, P.J., Rohn, K., Wolstenholme, A.J., 2009b. Molecular detection

- of benzimidazole resistance in *Haemonchus contortus* using real-time PCR and pyrosequencing. *Parasitol* 136, 349-358.
- Waghorn, G.C., Gregory, N.G., Todd, S.E., Wesselink, R., 1999. Dags in sheep; a look at faeces and reasons for dag formation. *Pr N Z Grassl Assoc* 61, 43-49.
- Walker, J., Hoekstra, R., Roos, M.H., Wiley, L.J., Weiss, A.S., Sangster, N.C., Tait, A., 2001. Cloning and structural analysis of partial acetylcholine receptor subunit genes from the parasitic nematode *Teladorsagia circumcincta*. *Vet Parasitol* 97, 329-335.
- Wang, C., Paul, M. 2017. eggCounts: Hierarchical Modelling of Faecal Egg Counts.
- Wang, C., Torgerson, P.R., Hoglund, J., Furrer, R., 2017. Zero-inflated hierarchical models for faecal egg counts to assess anthelmintic efficacy. *Vet Parasitol* 235, 20-28.
- Wang, T., van Wyk, J.A., Morrison, A., Morgan, E.R., 2014. Moisture requirements for the migration of *Haemonchus contortus* third stage larvae out of faeces. *Vet Parasitol* 204, 258-264.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2017. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol* 35, 543-548.
- Weetman, D., Wilding, C.S., Neafsey, D.E., Müller, P., Ochomo, E., Isaacs, A.T., Steen, K., Rippon, E.J., Morgan, J.C., Mawejje, H.D., Rigden, D.J., Okedi, L.M., Donnelly, M.J., 2018. Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African *Anopheles gambiae*. *Sci Rep* 8, 2920.
- Wenick, A.S., Hobert, O., 2004. Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in *C. elegans*. *Dev Cell* 6, 757-770.
- Wetterstrand, K. 2019. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).
- Wickham, H., 2007. Reshaping Data with the reshape Package. 2007 21, 20.
- Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York,.
- Williamson, S.M., Storey, B., Howell, S., Harper, K.M., Kaplan, R.M., Wolstenholme, A.J., 2011. Candidate anthelmintic resistance-associated gene expression and sequence polymorphisms in a triple-resistant field isolate of *Haemonchus contortus*. *Mol Biochem Parasitol* 180, 99-105.
- Williamson, S.M., Wolstenholme, A.J., 2012. P-glycoproteins of *Haemonchus contortus*: development of real-time PCR assays for gene expression studies. *J Helminthol* 86, 202-208.
- Wilson, D.J., Sargison, N.D., Scott, P.R., Penny, C.D., 2008. Epidemiology of gastrointestinal nematode parasitism in a commercial sheep flock and its implications for control programmes. *Vet Rec* 162, 546-550.
- Wimmer, B., Craig, B.H., Pilkington, J.G., Pemberton, J.M., 2004. Non-invasive assessment of parasitic nematode species diversity in wild Soay sheep using molecular markers. *Int J Parasitol* 34, 625-631.
- Winterrowd, C.A., Pomroy, W.E., Sangster, N.C., Johnson, S.S., Geary, T.G., 2003. Benzimidazole-resistant beta-tubulin alleles in a population of parasitic nematodes (*Cooperia oncophora*) of cattle. *Vet Parasitol* 117, 161-172.



- Wolstenholme, A.J., 2012. Glutamate-gated Chloride Channels. *J Biol Chem* 287, 40232-40238.
- Wong, K.H., Jin, Y., Moqtaderi, Z., 2013. Multiplex Illumina sequencing using DNA barcoding. *Curr Protoc Mol Biol* Chapter 7, Unit 7.11.
- Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15, R46.
- Wood, I.B., Amaral, N.K., Bairden, K., Duncan, J.L., Kassai, T., Malone Jr, J.B., Pankavich, J.A., Reinecke, R.K., Slocombe, O., Taylor, S.M., Vercruysse, J., 1995. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) second edition of guidelines for evaluating the efficacy of anthelmintics in ruminants (bovine, ovine, caprine). *Vet Parasitol* 58, 181-213.
- Xie, C., Tammi, M.T., 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* 10, 80.
- Xie, M., Roy, R., 2012. Increased Levels of Hydrogen Peroxide Induce a HIF-1-dependent Modification of Lipid Metabolism in AMPK Compromised *C. elegans* Dauer Larvae. *Cell Metab* 16, 322-335.
- Xu, M., Molento, M., Blackhall, W., Ribeiro, P., Beech, R., Prichard, R., 1998. Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog. *Mol Biochem Parasitol* 91, 327-335.
- Yadav, C.L., Grewal, H.S., Banerjee, D.P., 1993. Susceptibility of two crossbreeds of sheep to *Haemonchus contortus*. *Int J Parasitol* 23, 819-822.
- Yassin, L., Gillo, B., Kahan, T., Halevi, S., Eshel, M., Treinin, M., 2001. Characterization of the DEG-3/DES-2 Receptor: A Nicotinic Acetylcholine Receptor That Mutates to Cause Neuronal Degeneration. *Mol Cell Neurosci* 17, 589-599.
- Yemini, E., Jucikas, T., Grundy, L.J., Brown, A.E.X., Schafer, W.R., 2013. A database of *Caenorhabditis elegans* behavioral phenotypes. *Nat Methods* 10, 877.
- Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8, 28-36.
- Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., Belgrader, P., Price, A.D., Lowe, A.J., Marks, P., Vurens, G.M., Hardenbol, P., Montesclaros, L., Luo, M., Greenfield, L., Wong, A., Birch, D.E., Short, S.W., Bjornson, K.P., Patel, P., Hopmans, E.S., Wood, C., Kaur, S., Lockwood, G.K., Stafford, D., Delaney, J.P., Wu, I., Ordonez, H.S., Grimes, S.M., Greer, S., Lee, J.Y., Belhocine, K., Giorda, K.M., Heaton, W.H., McDermott, G.P., Bent, Z.W., Meschi, F., Kondov, N.O., Wilson, R., Bernate, J.A., Gauby, S., Kindwall, A., Bermejo, C., Fehr, A.N., Chan, A., Saxonov, S., Ness, K.D., Hindson, B.J., Ji, H.P., 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat Biotech* 34, 303.
- Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326-3328.
- Zhu, Y., Bergland, A.O., Gonzalez, J., Petrov, D.A., 2012. Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS One* 7, e41901.

Żyżyńska-Granica, B., Koziak, K., 2012. Identification of suitable reference genes for real-time PCR analysis of statin-treated human umbilical vein endothelial cells. *PLoS One* 7, e51547-e51547.