A major locus for ivermectin resistance in a parasitic nematode

Stephen R. Doyle[1]*, Christopher J. R. Illingworth[2]*, Roz Laing[3], David J. Bartley[4], Elizabeth Redman[5], Axel Martinelli[1+], Nancy Holroyd[1], Alison A. Morrison[4], Andrew Rezansoff[5], Alan Tracey[1], Eileen Devaney[3], Matthew Berriman[1], Neil Sargison[6], James A. Cotton[1**], John S. Gilleard[5**]

* Joint first authors

** Authors for correspondence

+ Current affiliation: Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, Sapporo, Japan, Biological and Environmental Sciences and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

1. Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom

Stephen R. Doyle: sd21@sanger.ac.uk

Alan Tracey: alt@sanger.ac.uk

James A. Cotton: jc17@sanger.ac.uk

Nancy Holroyd: neh@sanger.ac.uk

Matthew Berriman: mb4@sanger.ac.uk

Axel Martinelli: axel.martinelli@gmail.com

2. Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom

Chris J. R. Illingworth: cjri2@cam.ac.uk

(also Department of Applied Maths and Theoretical Physics, Wilberforce Road, Cambridge CB3 0WA)

3. Institute of Biodiversity Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Garscube Campus, Glasgow, G61 1QH, United Kingdom

Roz Laing: Rosalind.Laing@glasgow.ac.uk

Eileen Devaney: Eileen.Devaney@glasgow.ac.uk

4. Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik EH26 0PZ, United Kingdom

David J. Bartley: Dave.Bartley@moredun.ac.uk

Alison A. Morrison: Alison.Morrison@moredun.ac.uk

5. Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

John S. Gilleard: jsgillea@ucalgary.ca

Elizabeth Redman: elmredma@ucalgary.ca

Andrew Rezansoff: amrezans@ucalgary.ca

6. University of Edinburgh, Royal (Dick) School of Veterinary Studies, Edinburgh, EH25 9RG, United Kingdom

Neil Sargison: Neil.Sargison@ed.ac.uk

Abstract

Infections with helminths cause an enormous disease burden in humans, livestock, and companion animals. Helminth control is heavily dependent on anthelmintic drugs, but resistance to these drugs is widespread, and the genetic determinants of resistance remain unresolved. Here, we build on two previous genetic crosses between ivermectin resistant and sensitive isolates of the parasitic nematode *Haemonchus contortus*, an economically important gastrointestinal parasite of small ruminants and a model for anthelmintic research. By combining genetic mapping and genome resequencing with population genetic modelling, we identify the first single genomic locus conclusively associated with ivermectin resistance in a parasitic nematode. This locus is common between two geographically and genetically divergent strains and does not include the leading candidate genes. This work is the first example of a major anthelmintic resistance locus to be identified by genetic mapping and represents the most comprehensive genetic analysis of this trait in a parasitic helminth to date.

Introduction

Parasitic worms, commonly termed helminths, are extremely diverse and frequently responsible for significant morbidity and mortality in their hosts. Control of those of human and veterinary importance is heavily dependent on the large-scale administration of anthelmintic drugs; for instance, the macrocyclic lactone ivermectin has been extremely successful in the control of a number of helminths in both humans and animals (Omura and Crump, 2004). The importance of the discovery and development of this drug as an anthelmintic was recently recognised by the award of the 2015 Nobel Prize for Medicine or Physiology. However, these successes are now threatened by the emergence of drug resistance. In many species of parasitic nematodes of livestock, anthelmintic drug resistance is already a major problem for global agricultural production and animal welfare and its rapid acquisition has been documented against single and multiple drug classes (Kaplan and Vidyashankar, 2012; Rose et al., 2015). There are growing concerns regarding the reduced efficacy of compounds used in mass drug administration (MDA) programs for some human-infective helminths, which could drive the emergence of resistance (Crellen et al., 2016; Doyle et al., 2017; Geerts and Gryseels, 2000; Osei-Atweneboana et al., 2011). Consequently, anthelmintic drug resistance is an emerging threat to the sustainable control of human parasites, particularly those targeted by MDA programs. However, in spite of the importance of this issue, remarkably little is known regarding the molecular mechanisms of resistance to most anthelmintic drug groups, with the notable exception of the benzimidazole class. This is, at least in part, due to a lack of genomic resources, tools and techniques with which to study these experimentally challenging organisms.

Haemonchus contortus is a globally distributed parasite of wild and domesticated ruminants that has a major impact on the health and economic productivity of sheep and goats. Resistance of this parasite to almost all of the classes of anthelmintic drugs has been documented in multiple regions of the world (Yadav et al. 1995; Jabbar et al. 2013; Echevarria and Trindade 1989; Van den Brom et al. 2015; Howell et al. 2008). Isolates of *H. contortus* that are resistant to multiple classes simultaneously are common and resistance can occur within a just a few years of introduction of a new drug class (Williamson et al., 2011; van Wyk and Malan, 1988). Partly for these reasons, *H. contortus* has emerged as a model parasitic nematode for anthelmintic resistance as well as drug and vaccine discovery research (Geary, 2016; Gilleard, 2013; Nisbet et al., 2016). Its utility as a model is largely due to a greater amenability to experimentation than most parasitic nematodes; it is possible to establish and maintain isolates *in vivo* in the natural host, perform genetic crosses *in vivo*, and undertake *in vitro* culture for at least part of its life cycle, allowing drug assays and genetic manipulation such as RNAi to be performed (Britton et al., 2016). The ability to utilise these molecular approaches is complemented by extensive information about the structure of the genome and transcriptional differentiation of the major life stages (Doyle et al., 2018; Laing et al., 2013; Schwarz et al., 2013).

Genetic characterisation of drug resistance in parasitic helminths will enable sensitive and specific tools to be developed for research, surveillance and clinical diagnostics to support sustainable parasite management programs. Understanding the mechanisms of resistance to current drugs should also provide important

information to help direct the development of new drugs (Hefnawy et al. 2017; Cowell et al. 2018). However, the identification of genetic loci associated with drug resistance in these organisms remains a significant challenge. The only anthelmintic drug class to date for which the major resistance loci have been unequivocally identified is the benzimidazoles, where three mutations in the the isotype-1 ß-tubulin gene of *H. contortus* and several other trichostrongylid nematodes species are the major causal loci (Elard and Humbert, 1999; Kwa et al., 1994; Rufener et al., 2009; Silvestre and Cabaret, 2002). In the case of other anthelmintic drug classes, the genetic mediators of resistance remain obscure; for example, mutagenesis experiments in the free-living nematode *Caenorhabditis elegans* have identified many genetic loci capable of conferring low levels of ivermectin resistance, however, very few mutations confer high levels of resistance (Hunt, 1995). Concurrent mutation of three genes, *Ce-avr-14*, *Ce-avr-15* and *Ce-glc-1,* each encoding a glutamate-gated chloride channel (GluCl), confers a high level of resistance to ivermectin (Dent et al., 2000). However, mutations in orthologous GluCl genes have not been associated with ivermectin resistance in parasitic nematodes. In the case of parasitic helminths, largely due to the lack of genomic resources, research into the genetic basis of ivermectin resistance has been dominated by the examination of small numbers of candidates genes (Kotze et al., 2014). Based on their potential roles in the mechanism of action or efflux of drugs, candidate genes are generally examined for associations between sequence polymorphisms and resistance phenotypes. However, the extremely high levels of genetic diversity of *H. contortus,* together with the limited number of well characterised ivermectin resistant isolates, has provided a major challenge to such studies. The overall outcome of such

candidate gene studies to date has provided at best circumstantial and inconsistent support for the involvement of any of the leading candidate genes in ivermectin resistance. It has also led to substantial debate regarding the extent to which ivermectin resistance is polygenic with the phenotype being primarily due to the additive effects of a large number of minor effect loci.

Genome-wide and genetic mapping approaches are emerging for *H. contortus* due to progress in genetic crossing methodologies (Sargison et al., 2018) and an increasing complement of genomic resources. Two serial backcrosses have previously been undertaken between the genome reference susceptible strain MHco3(ISE) and two geographically independent ivermectin resistant isolates, MHco4(WRS) and MHco10(CAVR) (Redman et al., 2012) (**Figure 1 A,B**). Using a panel of microsatellite markers, the marker Hcms8a20 confirmed that these backcrosses had successfully introgressed ivermectin resistance loci from the two resistant strains into the susceptible MHco3(ISE) genomic background. Subsequent genetic studies showed that none of the leading candidate genes from the literature - *Hco-avr-14, Hco-glc-5, Hco-lgc-37, Hco-pgp-9, Hco-pgp-2* and *Hco-dyf-7* - showed evidence of introgression (Rezansoff et al., 2016). The increasing accessibility to high throughput sequencing of non-model organisms, together with an increasingly high quality reference for *H. contortus*, offers an opportunity to characterise precisely the genome-wide evidence of introgression and selection surrounding the Hcms8a20 locus.

Here we build upon these two previous genetic crosses, extending them for further generations of passage with ivermectin selection (**Figure 1 C**), and generating whole genome sequencing data from four steps in the experiment. We analyse these data with traditional population genetic and novel statistical methods, and exploit a chromosome-scale genome assembly to map ivermectin resistance in the two genetically and geographically distinct strains of *H. contortus*. Previous research has highlighted the value of novel methods in population genetics for analysing experimental cross populations (Illingworth and Mustonen, 2013). In order to gain the maximum possible insight into the data collected, we extended previous statistical work (Abkallo et al., 2017; Illingworth et al., 2012) to account for stochasticity in the experiment. Specifically, the random location of recombination events, in the context of strong selection, and genetic drift imposed by population bottlenecks in the experimental design, lead to potential variation in the final outcome of the experiment. With prior knowledge of the recombination rate (Doyle et al. 2018), we used evolutionary simulations reproducing the experimental design to explore the range of experimental outcomes arising under multiple scenarios of selection; this allowed for a direct inference of evolutionary parameters. Together, our methods identify a single region of genomic introgression under selection for resistance. This work represents the most comprehensive analysis of the genetics of anthelmintic resistance in a parasitic nematode to date. Further, the study demonstrates the power of using a genetic crossing approach, in the context of a contiguous genome assembly, to eliminate false positive genetic signals typically linked to resistance.

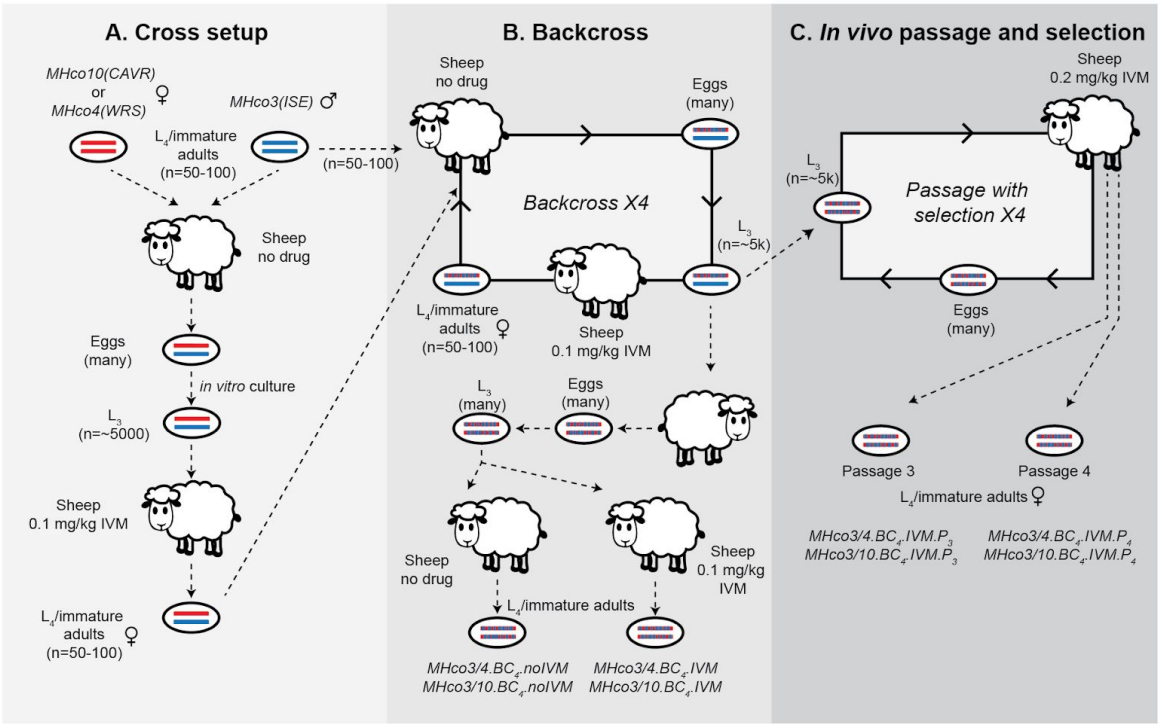


Figure 1. Outline of the initial crosses, backcrosses, *in vivo* passage and selection experiments. A. Ivermectin resistant strains MHco10(CAVR) or MHco4(WRS) ("resistant" haplotypes are depicted as red lines) were crossed with an ivermectin sensitive strain MHco3(ISE) ("susceptible" haplotypes as blue lines) to generate heterozygous F_1 progeny. F_1 eggs were cultured *in vitro* to L_3, before maturation *in vivo* to L_4/immature adults, from which females were used to initiate the backcross. **B.** The first round of backcross was performed by crossing heterozygous females from the initial cross with susceptible MHco3(ISE) males *in vivo*, resulting in F_2 progeny with reduced heterozygosity due to enrichment of MHco3(ISE) haplotypes. Resistance alleles were maintained in the backcross population by selection with ivermectin, before seeding the next round of the backcross with cross-derived heterozygous females and new susceptible MHco3(ISE) males. This process was repeated for four rounds of backcrossing, resulting in the backcrossed population becoming genetically similar to the MHco3(ISE) parental line in all regions of the genome not linked to ivermectin resistance. After four rounds of backcrossing, introgressed L_3 progeny were used to infect a new recipient sheep, resulting in segregation of susceptible and resistant alleles in both haplotypes among the progeny (mixed red/blue haplotypes). Eggs were cultured to L_3 before infecting two sheep, one exposed to ivermectin and one that did not receive drug. Post ivermectin treatment, L_4/immature adults from both sheep were recovered on necropsy for sequencing. **C.** Post-backcross L_3 were further passaged into a worm-naive sheep and treated with ivermectin, after which eggs were recovered and cultured to infective L_3 for reinfection. This process was repeated for four generations of

passage with selection (but without backcrossing). L_4/immature adults were recovered after passage three and four for sequence analysis.

Figure Supplement

Table S1. Sequencing datasets used in this study

Results

Genetic diversity within and between parental *H. contortus* strains

Genome-wide genetic diversity was characterised within and between two ivermectin resistant strains (MHco10[CAVR] and MHco4[WRS]) and one ivermectin susceptible strain (MHco3[ISE]) of *H. contortus*. We exploited a greatly improved, chromosomal-scale *H. contortus* reference genome sequence (**Figure 2A**) that was derived from an inbred isolate of the MHco3(ISE) strain. Whole genome sequencing of pooled worms from each of the geographically distinct strains (**Figure 2B**) revealed high levels of nucleotide diversity throughout the five autosomes of the genome, with the average diversity almost twice as high in the resistant strains (MHco10[CAVR] mean π = 0.035 ± 0.008 standard deviations (SD); MHco4[WRS] mean π = 0.038 ± 0.008 SD) than the susceptible one (MHco3[ISE] mean π = 0.022 ± 0.008 SD) (**Figure 2C; Figure S1A**). In contrast to the autosomes, the X chromosomal scaffolds were significantly less diverse in all three strains analysed (MHco3[ISE] mean π = 0.008 ± 0.008, MHco10[CAVR] mean π = 0.014 ± 0.012, MHco4[WRS] mean π = 0.017 ± 0.014) (**Figure 2C; Figure S1A**), consistent with our recent finding that the X chromosome contains as little as 10% as much genetic diversity relative to the autosomes (Doyle et al., 2018). The difference between sensitive and resistant strains is perhaps not surprising given their history; MHco3(ISE) was subject to multiple rounds for inbreeding during is original derivation as a laboratory strain (Roos et al. 2004), whereas the resistant strains have not been subjected to experimental inbreeding and were isolated from outbred populations more recently.

We examined short-range haplotype diversity in each strain by measuring linkage disequilibrium (LD) between pairs of SNPs detected in paired reads (**Figure S1B**). Although restricted by the Pool-seq design to pairwise SNP comparisons less than 500 bp apart, i.e., within a read pair, we nevertheless observed considerable decay in LD over this distance. Moreover, the rate of LD decay was correlated with nucleotide diversity, such that there was a greater loss of LD in the MHco4(WRS) and MHco10(CAVR) strains over the 500 bp, and a greater proportion of variants in LD in the less diverse, susceptible MHco3(ISE) strain.

To understand the extent of genetic diversity that is private to, and thus unlikely to be shared between, each population, we identified variant positions in the genome whereby the variant frequency was enriched in one population but not the other two populations. Of the ~7.6 million biallelic SNPs distributed in the genomes of the samples analysed, ~514 k were private to MHco3(ISE) (**Figure S2A**), ~960 k to MHco10(CAVR) (**Figure S2B**) and ~685 k to MHco4(WRS) (**Figure S2C**). Each strain contained local regions of high diversity that differentiated it from the others.

We also explored larger structural variation within each of the strains, including putative deletions, duplications and inversions detected by the presence of discordant and split read pairs among the perfectly paired reads (**Figure S3**). We note that structural variant calling in pooled individual, short read data will be insensitive and subject to false positive calls due to mismapping. However, even a conservative approach examining only high quality homozygous variants revealed significant variation between strains (**Figure S3 A,C,E**), with a large number of

deletions, and substantially fewer duplications and inversions detected (**Figure S3 B,D,F**).

Genetic diversity between each of the three strains was assessed by pairwise analysis (measured by F_{ST} of single nucleotide polymorphisms [SNPs] in 10 kb windows), which confirmed significant genome-wide strain differentiation (**Figure 2D**). Most of this genetic differentiation reflects underlying genetic structure between strains, due to a degree of reproductive isolation due to their geographic distribution (**Figure 2B**). Collectively, these data emphasise the challenge of characterising genetic variation associated with phenotypes such as drug resistance by simply comparing genetic diversity between a susceptible and resistant strain without accounting for this background genetic variation.

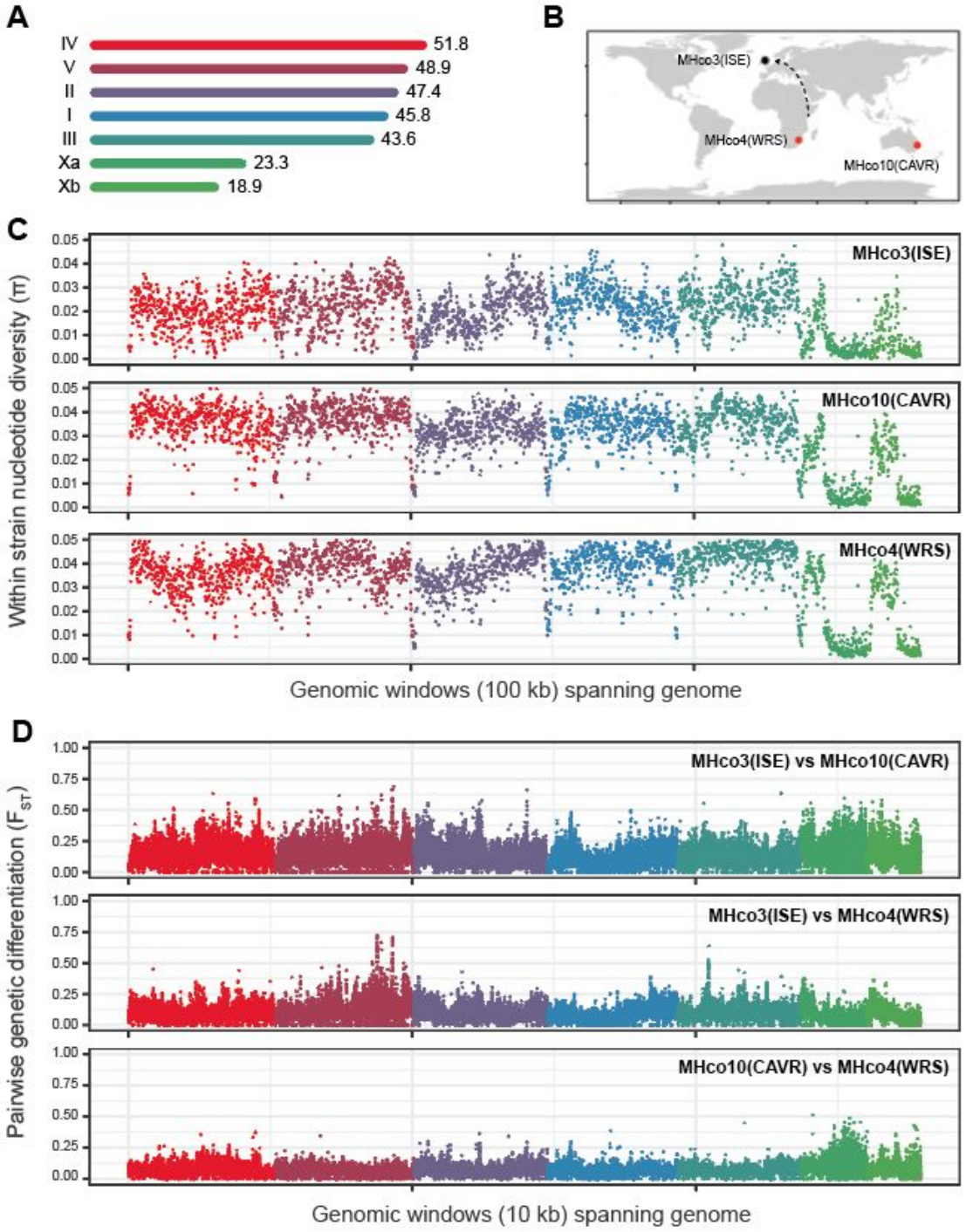


Figure 2. Within and between strain genetic diversity of parental isolates used in the genetic cross. A. The 279 Mb genome assembly of *H. contortus* consists of five autosomal and two X chromosome scaffolds, named based on synteny with *Caenorhabditis elegans* chromosomes. The size of each scaffold is indicated, and are presented in order by length (Mb). **B.** Geographic origin of the susceptible MHco3(ISE) and ivermectin resistant MHco10(CAVR) and MHco4(WRS) strains used in the genetic crosses. All strains are archived at the Moredun Research Institute, UK – while MHco3(ISE) has been maintained for decades, it was originally

isolated in East Africa. **C.** Within-strain nucleotide diversity for each of the parental strains, calculated as mean diversity per 100 kb windows throughout the genome using *npstat*. **D.** Between strain diversity, calculated as pairwise F_{ST} in 10 kb windows throughout the genome using *popoolation2*. Colours here and throughout represent chromosomes as described in A.

Figure supplements

Figure S1. Characterisation of within-strain diversity.

Figure S2. Distribution of "private" variant sites per parental population.

Figure S3. Copy number and structural variation in the parental populations.

Genome-wide analysis of genetic diversity reveals the same single large introgressed region in both backcross lines

A backcross has previously been performed to introgress resistance-conferring alleles from the MHco10(CAVR) and MHco4(WRS) strains into the MHco3(ISE) susceptible genetic background (Redman et al., 2012). In the present study, we used whole genome pairwise comparisons of the genetic diversity throughout the backcross to assess the degree of introgression and to detect evidence of selection associated with resistance. Each pairwise comparison was between particular generations of each backcross and the MHco3(ISE) susceptible parent. We hypothesised that where backcross populations were not subjected to further ivermectin treatment (Figure 1: MHco3/10.BC_4.noIVM & MHco3/4.BC_4.noIVM), there would be relatively little genetic differentiation across the majority of the genome between the population and MHco3(ISE). By contrast, where populations were subjected to further ivermectin treatment, we hypothesised that genetic differentiation with MHco3(ISE) should increase close to any resistant allele, with

high levels of differentiation in regions of the genome linked to an ivermectin resistance-conferring locus and low levels of differentiation elsewhere.

As hypothesised in both crosses, there was relatively little genetic differentiation between the post-backcross, no-selection population and the susceptible parental strain (**Figure 3**; MHco3/10.BC_4.noIVM and MHco3/4.BC_4.noIVM). After backcrossing, in the absence of selection, genetic material from the susceptible parent should comprise approximately 97% of the population due to the repeated backcrossing with MHco3(ISE) (resistant alleles comprise half of the initial cross population and reduce by a half with each backcross), making them largely indistinguishable from the susceptible parent. Selection increases this but is limited by the backcross; no more than 50% of the material at any location may originate from the resistant parent. Upon further selection (**Figure 3**; MHco3/10.BC_4.IVM and MHco3/4.BC_4.IVM), a region of differentiation from the susceptible parent, particularly in the MHco3/10 cross, was localised to the right arm of chromosome V. This region was markedly differentiated in both MHco3/10 and MHco3/4 crosses after *in-vivo* selection and passage (**Figure 3**; MHco3/10.BC_4.IVM.P_3, MHco3/4.BC_4.IVM.P_3, MHc3/10.BC_4.IVM.P_4 and MHc3/10.BC_4.IVM.P_4). No other region of the genome displayed a marked and progressive increase in differentiation throughout the crosses. To test this explicitly, we determined the correlation between change in F_{ST} per each genomic window and progression through the cross; only the chromosome V region demonstrated progressive genetic differentiation over time (**Figure S4 A,B**), which was highly correlated between the two crosses (**Figure S4C**).

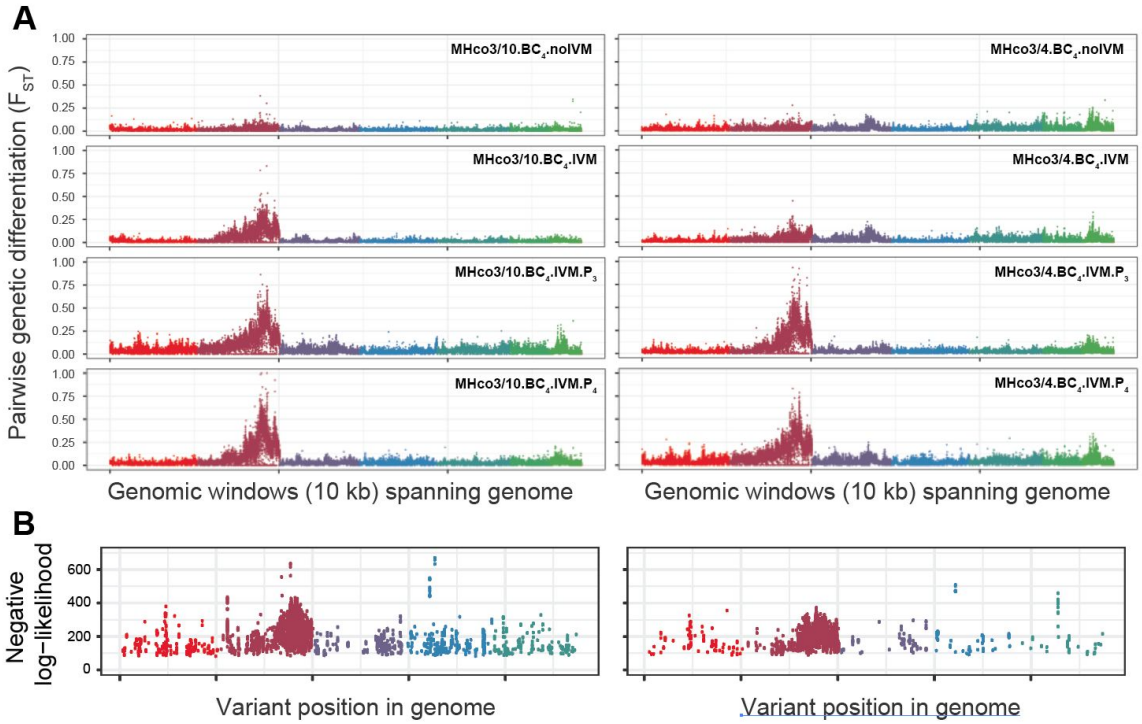


Figure 3. Genome-wide analysis of ivermectin associated loci. A. Pairwise genetic diversity throughout the MHco3/10 (left plots) and MHco3/4 (right plots) backcrosses. In both crosses and at each point in the backcross, the experimental strain was backcrossed against the MHco3(ISE) parental strain – the data is therefore presented to compare the genetic diversity (F_{ST} measure in 10 kb windows using *popoolation2*) between MHco3(ISE) and each sampled time point in the cross as per the cross schema in Figure 1. **B**. Output of the single-locus evolutionary model, describing sites which were inconsistent with a model of neutral evolution, measured using a likelihood threshold. Note that the X chromosome is not represented. Colours represent chromosomes as described in Figure 2A.

Figure supplements

Figure S4. Summary of genome-wide change in F_{ST} throughout the backcross and subsequent passage.

Figure S5. Fits between the model and the data for each data sample.

Figure S6. Location of significantly non-neutral loci identified using the single-locus population genetic model.

Collectively, these data suggest that there is a single introgression region on chromosome V that is under selection upon drug exposure, and that this region is under selection in the two independent serial backcrosses performed. This conclusion was supported by an independent evolutionary analysis, which compared the observed frequencies of variants associated with the resistant parental population with the distribution of allele frequencies expected under the assumption of selective neutrality (**Figure S6**). An over-representation of alleles with atypical frequencies further suggests that a single region in the parasite genome was selected towards the right hand end of chromosome V. In this region, the allele frequencies were strongly inconsistent with the neutral expectation, suggesting that they had potentially evolved under the influence of selection (**Figure 3B**), with 71% and 91% of the total number of such alleles for the MHco3/10 (**Figure S6 A)** and MHco3/4 (**Figure S6 B)** crosses, respectively.

Detailed analysis of introgression region on chromosome 5

To more precisely characterise the introgression region, we focus on chromosome V (**Figure 4**). A comparison of genetic diversity between MHco3(ISE) and the end-point of both crosses (BC_4.IVM.P_4) revealed a strikingly similar pattern between the crosses, with a major region of differentiation between 38-42 Mb, and a second but lesser increase in genetic differentiation between 45-48 Mb that is most prominent in the MHco3/10 (**Figure 4A**). Encouragingly, the major region of introgression lies immediately adjacent to the Hcms8a20 microsatellite marker (located at position 36.16 Mb along chromosome V) that was shown to be genetically

linked to ivermectin resistance in the preliminary analysis of these backcrosses (Redman et al., 2012).

The F_{ST} data are supported by a comparison of Tajima's D throughout this region (**Figure 4B**; top plots). Tajima's D test statistic is commonly used to distinguish neutrally evolving loci (D ~ 0) from non-neutral processes, such as balancing (D > 0) or positive selection (D < 0). Although variation in Tajima's D was observed throughout the genome in both crosses (**Figure S7**), the greatest variation between the susceptible parent and backcrossed lines was observed at approximately 39-40 Mb (**Figure 4B**; bottom plots), which was associated with significantly negative Tajima's D in the backcrossed lines consistent with strong positive selection. In contrast to the F_{ST} data, analysis using Tajima's D suggested a single peak, corresponding to a single site or region under positive selection.

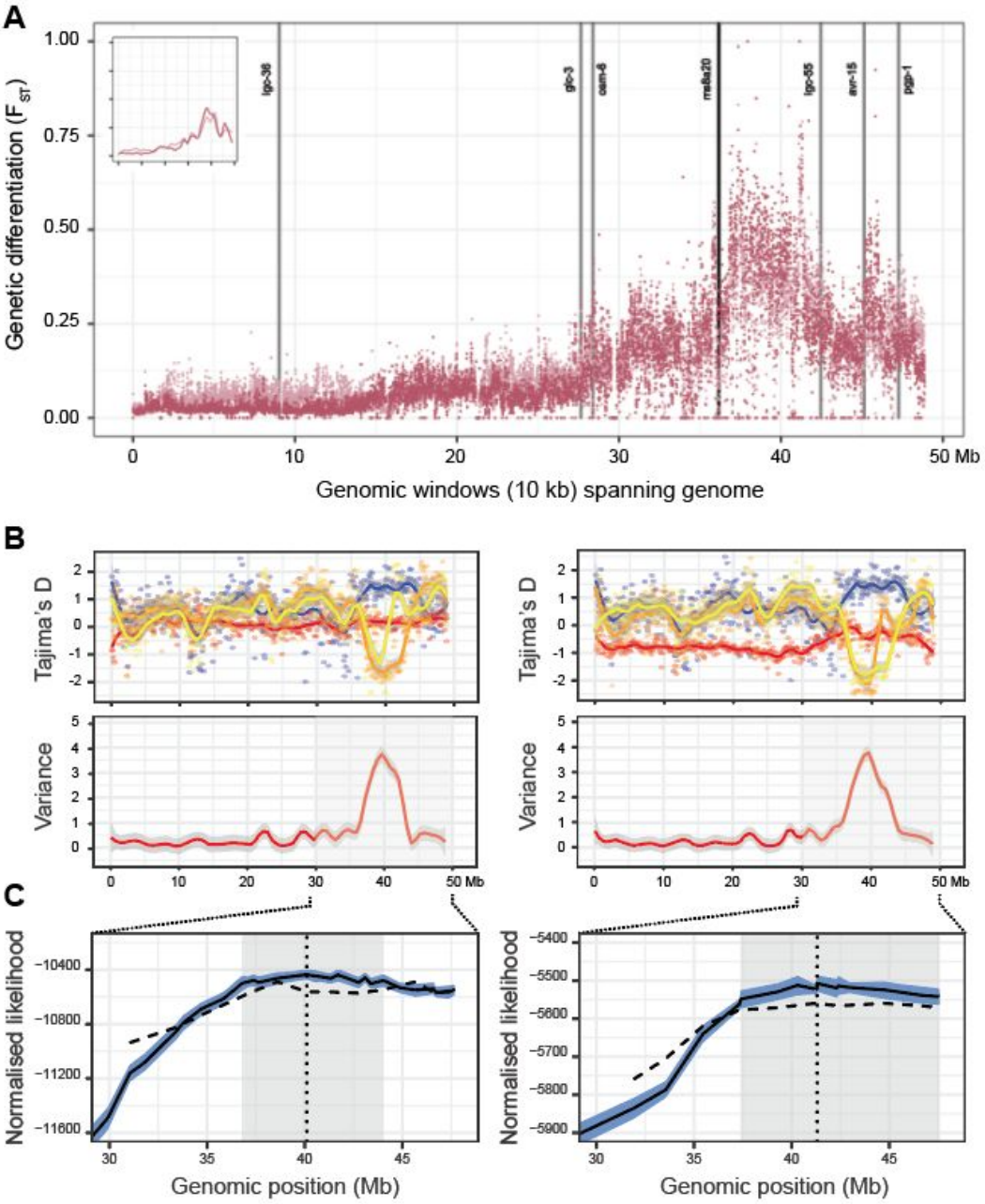


Figure 4. Analysis of chromosome V and the major region of introgression. A. Genetic differentiation (F_{ST}) measured in 10 kb windows throughout chromosome V is presented between MHco3(ISE) parent and both MHco3/10.BC_4.IVM.P_4 (light) and MHco3/4.BC_4.IVM.P_4 (dark) representing the final time point of the crosses. Inset presents the smoothed F_{ST} distribution of the two comparisons. Published candidate genes in or near the introgression region (grey vertical lines), as well as the original Hcms8a20 microsatellite marker that Redman *et al.* (2012) linked to ivermectin resistance (black vertical line), are presented. **B.** Comparison of Tajima's D per chromosome between MHco3(ISE) parent (blue), MHco10(CAVR) (left panels) or MHco4(WRS) (right panels) and passage 3 (MHco3/10.BC_4.IVM.P_3 or MHco3/4.BC_4.IVM.P_3; orange) and passage 4 (MHco3/10.BC_4.IVM.P_4 or

MHco3/4.BC_4.IVM.P_4; yellow) of the crosses. Tajima's D was calculated using *npstat* in 100 kb windows spanning the genome. The variance in the mean value of Tajima's D among MHco3(ISE) and passages 3 and 4 samples is presented as smoothed line (red). The null expectation is that variance between these samples will be low in regions of the genome not under selection. **C**. Inferred likelihoods from the multi-locus population genetic model. The mean likelihood of the best fitting single-locus model at each locus position is shown by the solid black line; the light blue interval around this line shows the 5-95% confidence interval for this statistic calculated by a bootstrapping process. The position of the maximum likelihood value of the likelihood is shown by the vertical black dotted line; a confidence interval for this position, calculated from the bootstrapping values, is shaded in gray. The mean likelihood of the best fitting two-locus model at including each locus position is shown by the black dashed line.

Figure supplements

Figure S7. Analysis of Tajima's D variation in each chromosome per cross.

Figure S8. Contour maps of log likelihood scores derived from the two locus driver model.

Figure S9. Relative position and location of candidate genes from literature proposed to be associated with ivermectin resistance in *Haemonchus contortus* and / or *Caenorhabditis elegans.*

Table S1. Coordinates of candidate genes from literature proposed to be associated with ivermectin resistance in *Haemonchus contortus* and/or *Caenorhabditis elegans*

In order to investigate whether one or more than one selected allele caused the observed evolutionary signal, a multi-locus population genetic model was applied. In contrast to Fst and Tajima's D, this approach explicitly considers linkage between variant alleles arising from their common parental origin. This approach identified a location for the selected resistance-conferring allele that was consistent with the previous analysis. In the first step of the model, putative segregating sites in chromosome V between the resistant and susceptible parental populations were identified; a total of 474 such sites were identified in the MHco3/10 dataset, with 157 sites in the MHco3/4 dataset; calculations were performed upon data from these loci. A maximum likelihood model inferred the location of the driver to be in a broadly

consistent location in each experiment, with peaks at 40.10 and 41.31 Mb respectively. However, there was considerable uncertainty in the location of the selected allele, with confidence intervals of 36.86 to 44.03 Mb and 37.41 to 47.56 Mb in each case (**Figure 4 C**). These broad confidence intervals were explained by the population structures inferred by the model, which explicitly considered the location of crossover recombination events between the parental strains throughout the experiment. While the precise location of crossover events is stochastic, repeated calculations produced an ensemble of replicate population structures for each set of model parameters. Example structures generated using the maximum likelihood parameters for each population show how selection for resistance drives the accumulation of genetic material from the resistant parent through the course of the experiment (**Figure 5**); these outputs may be contrasted with equivalent data generated under a model of selective neutrality (**Figure S10**). Across 250 replicates, a mean of 177.1 crossover recombination events were predicted in chromosome V of the final MHco3/10 population (range 120 to 224), giving a mean length of a block of parental genome of 17.64 Mb; in the final MHco3/4 population a mean of 186.4 recombination events were seen (range 143 to 238), giving a mean length of a block of parental genome of 17.06 Mb. Under selection, blocks of genome containing the selected allele are favoured; the large size of each block reduces the precision with which the location of the allele under selection can be identified.

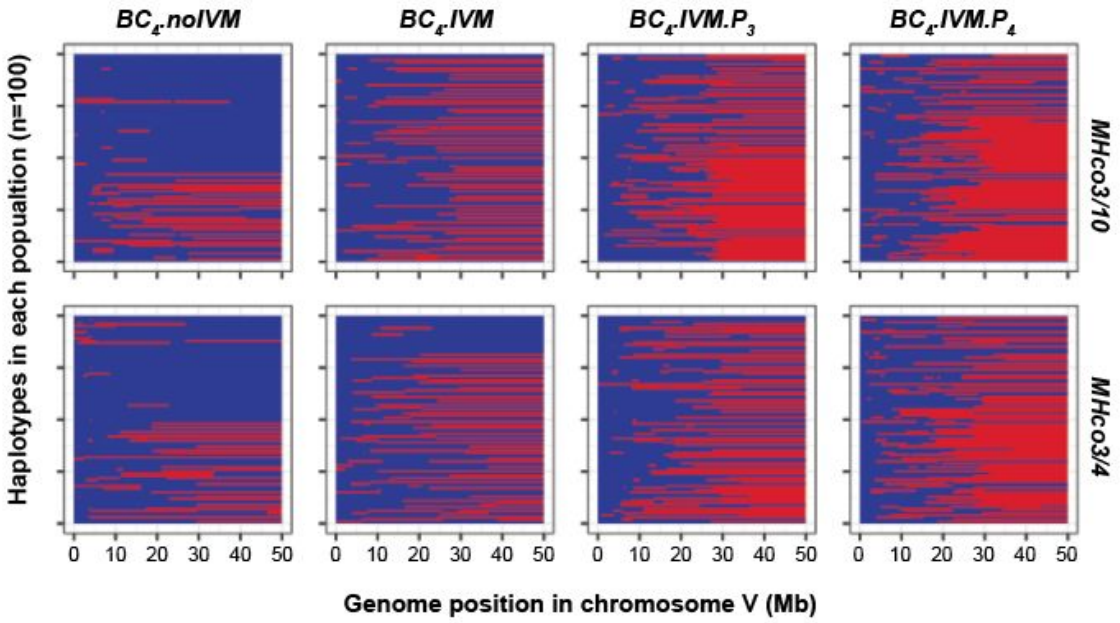


Figure 5. Haplotype structure of chromosome V in an example output from the multi-locus model. Segments of genome from the parent containing the resistance allele are shown in red, while segments of genome from the susceptible parent are shown in blue. Data shown were generated using the maximum likelihood parameters in each case.

Figure supplements

Figure S10. Haplotype structure of chromosome V in an example output from the two driver model under neutral evolution.

Our analysis suggests that the limited extent of recombination in the population, rather than selection for a second resistance allele, may explain the output of the experiment. An extended multi-locus model incorporating selection on alleles at two loci within the introgression region was used, investigating the potential for a second driver to explain the second peak of differentiation found at approximately 45 Mb in the F_{ST} analysis. Some separation of selected alleles by recombination is required for these alleles to be distinguished; we therefore applied the constraint that the selected alleles must be at least 2 Mb apart in the genome. Under this constraint, the

best fitting model identified variants under selection at the positions 38.6 Mb and 45.7 Mb in the MHco3/10 data (**Figure S8 A)**, and at the positions 41.2 Mb and 44.8 Mb in the MHco3/4 data (**Figure S8 B**). However, under the imposed constraint, the two-driver model performed worse than the single-driver model shown above, favouring an explanation in which only one variant is under selection. (We note that removing the constraint that selected alleles be separated would guarantee a likelihood at least equal to that of the single-driver model). In conclusion, due to the limited extent of recombination in the experiment, we could not rule out the presence of more than one drug-resistance variant in the identified region of chromosome V, but we did not find evidence in favour of a second selected allele.

Many candidate genes have been proposed in the literature as being in association with ivermectin resistance. We curated a list of genes that have been explored in *H. contortus*, and/or had been shown to confer ivermectin resistance when mutated in *C. elegans*. The current genome assembly is not annotated, so we determined their location, either via mapping the published gene models from the V1 annotation, or determining the closest *H. contortus* orthologous gene from *C. elegans* candidates using Wormbase Parasite (Howe et al., 2017) (**Table S2; Figure S9**). At least one candidate gene was located in each chromosome; in chromosome V, the location of six candidate genes were determined. However, none of these genes were found in the main introgression region defined by the F_{ST} analysis (**Figure 4 A**; vertical annotated lines). Three genes - *Hco-lgc-55, Hco-avr-15,* and *Hco-pgp-1(9)*, the latter two which have a strong association with ivermectin resistance - lie to the telomere-side of the introgression region, however, they are located on the periphery

of the F_{ST} peaks rather than within them, suggesting that they are unlikely candidates to be under direct selection. Considering the absence of candidate resistance genes in the F_{ST} peaks, we conclude that the driver(s) of ivermectin resistance in the MHco4(CAVR) and MHco4(WRS) strains are novel and have not been previously described in association with ivermectin resistance.

Strength of selection

We further used the single locus model to infer the strength and manner of selection in favour of the drug resistance allele. Data in each case indicated strong selection for the resistant allele (**Figure 6**), such that susceptible worms produce multiple times fewer offspring than the resistant worms under drug treatment. In each case the value of the dominance coefficient, *h,* was less than one, indicating either an additive effect for the MHco10(CAVR) case, whereby each copy of the resistance allele contributes an equal amount of resistance, or of weak dominance for the MHco4(WRS) data, whereby the first resistance allele contributes slightly more than the second. This is consistent with the phenotypic characterisation of the initial crosses, whereby the F_1 individuals that would have been heterozygous for resistant alleles were partially resistant to ivermectin (Redman et al., 2012). As such, having two copies of the resistant allele had a greater phenotypic effect than one; these data therefore have implications for the mechanism of the drug resistance allele, and evolution of drug resistance in general terms, as discussed further below.

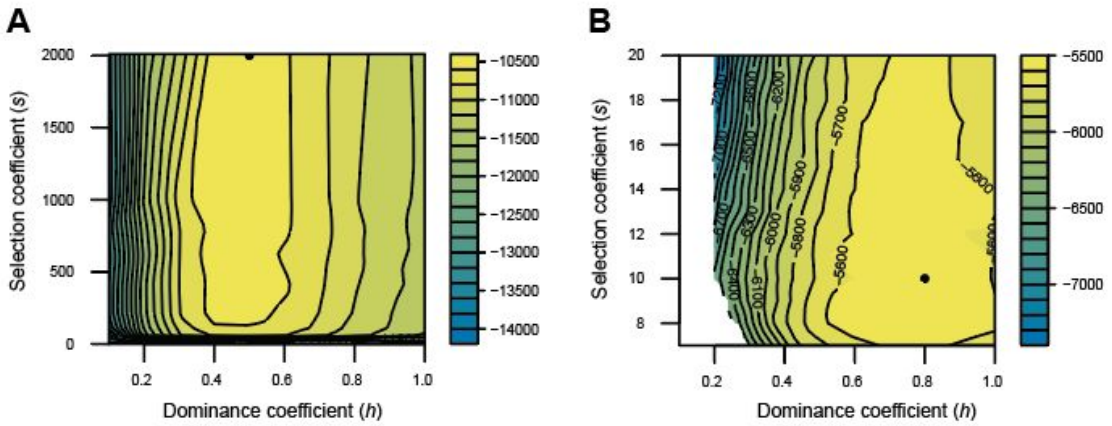


Figure 6: Likelihood surface describing inferred selection parameters for the resistance allele. Likelihoods in each landscape are reported for the case in which the position in the genome of the resistant allele was at the maximum likelihood location for both the MHco10(CAVR) (**A**) and MHco4(WRS) (**B**) parental strains. Within our model the fitness of a genotype is determined by the alleles at a single resistant allele, with homozygous susceptible worms having fitness 1, homozygous resistant worms having fitness 1+*s*, and heterozygous worms having fitness 1+*hs* in the presence of the drug. The black point in each plot gives the location of the maximum likelihood values for selection.

Discussion

Genetic mapping identifies the same major ivermectin resistance locus in two independant *H. contortus* strains from different continents

Given the importance of ivermectin, and other macrocyclic lactones, to human and animal health, the widespread emergence of resistance in multiple parasite species is alarming. In spite of several decades of research, the molecular genetic basis of ivermectin resistance in parasitic nematodes has remained elusive. This is due to a combination of factors including a lack of genomic resources, the general experimental intractability of parasitic nematodes and the often high level of genetic variation that complicates genetic analysis. *H. contortus* is the most experimentally tractable and most widely accepted parasite model for anthelmintic resistance research (Gilleard, 2013). Many candidate genes have been proposed to be associated with ivermectin resistance in *H. contortus*, chosen largely on *a priori* assumptions regarding their known or predicted function. Such candidate genes have typically been implicated as involved in ivermectin resistance largely on the basis of studies comparing SNP or haplotype frequencies between small numbers of resistance and susceptible isolates (Blackhall et al. 1998; Eng et al. 2006; Luo et al. 2017; Blackhall et al. 2003; Urdaneta-Marquez et al. 2014) or pre- and post-ivermectin treatment (Raza et al. 2016; Lloberas et al. 2013). However, careful validation of a number of these candidates in different field populations (Laing et al., 2016) or in controlled crosses (Rezansoff et al., 2016) have failed to support the initial associations observed. The discordance among studies reflects the difficulty in identifying real associations between genotype and phenotype simply by comparing resistant and susceptible populations, or individuals, due to the confounding effects

of extremely high levels of genome-wide genetic variation. The lack of consistency between studies had led to much discussion and debate regarding the complexity of the genetic basis of ivermectin resistance both in terms of the number of loci involved and the extent of geographical variation (Kotze et al. 2014; Prichard 2001; Beech et al. 2011). We have undertaken a genetic mapping approach to systematically identify the major ivermectin loci, to help resolve the lack of clarity and debate.

We have analysed whole genome sequence data collected from the most recent phase of a multi-generational crossing experiment conducted in populations of the parasitic worm *H. contortus*. Our analysis unequivocally identified a single region within chromosome V that contains at least one major ivermectin resistance allele. This major ivermectin resistance locus is common to two independent ivermectin strains - MHco10(CAVR) and MHco4(WRS) - that were originally isolated from Australia and Africa, respectively. This genomic region was clearly distinguished in the experimental data using different population genetic methods. Although this region in which we infer a selected allele to exist is relatively large, each of our analytical approaches consistently identified the same region under selection, with no other comparable regions of introgression elsewhere in the genome in either of the two backcross experiments. This suggests this introgression region is the most important ivermectin resistance locus in the MHco10(CAVR) and MHco4(WRS) strains.

Although there has been much discussion regarding the likely multigenic basis of ivermectin resistance, these results strongly suggest a single locus (or potentially multiple closely linked loci) is likely to be the major effector of ivermectin resistance in these two strains. We acknowledge that we cannot discount previously described candidate genes as mediators of resistance in other isolates of *H. contortus*, or in other nematode species. Moreover, our data does not exclude a driver of transcriptional regulation within the introgression region that is under selection, which itself influences expression of other genes, including candidate genes, outside of the introgression region. However, most of the leading published candidate ivermectin resistance genes including *Hco-glc-5*, *Hco-avr-14*, *Hco-pgp-9*, *Hco-lgc-37*, *Hco-pgp-2* and *Hco-dyf-7* were not located in or near the region of introgression, which is consistent with a recent study using targeted sequencing of these genes in which none of them showed a signal of introgression in these two backcross experiments (Rezansoff 2016). This suggests that none of these candidate genes are a major ivermectin resistance-conferring locus in the MHco10(CAVR) and MHco4(WRS) strains. Although we identified three other previously described candidate genes - *Hco-lgc-55*, *Hco-avr-15*, and *Hco-pgp-1* - adjacent to the introgression region, none were found within the peak of the region suggesting that these also are unlikely to be major direct targets of ivermectin-mediated selection.

What might account for the large number of candidate genes previously suggested to be associated with ivermectin resistance? The answer might lie in our direct comparison of genetic diversity between the parental strains (**Figure 2D**). This visualisation of the extent of genetic divergence between these strains makes it

easier to understand how particular sequence polymorphisms might be naively attributed to being associated with resistance when in fact they simply represent one of many strain specific genetic variations that occur as a result of the independent evolutionary histories and lack of interbreeding between the strains. Our results highlight the challenge of interpreting simple direct comparison of strains, even with genomic approaches, when trying to disentangle those genetic polymorphisms underlying resistance from a complex background of genetic variation independent of a strains resistance status. The use of controlled genetic crosses as presented here and elsewhere (Choi et al., 2017; Redman et al., 2012; Rezansoff et al., 2016), whereby population genetic structure can be explicitly accounted for in the analysis, provides a powerful way to mitigate the challenge of discriminating resistance-causing alleles from background genetic variation. As such, the locus we have identified on chromosome V represents the only unequivocal major ivermectin resistance locus identified in any parasitic nematode species to date.

Population genetic and evolutionary modelling defined the boundaries of the ivermectin resistance conferring locus

Inference of selection from the multi-locus model suggested the presence of stronger selection for the drug resistance allele in the MHco3/10 cross than in the MHco3/4 cross (**Figure 6**). Additionally, although the data suggests that resistance is at least additive in both crosses, i.e., heterozygotes are likely to confer some resistance, but increased resistance is achieved by having two resistant alleles, there is some evidence for dominance in the MHco4(WRS) strain, whereby the second resistant allele confers a lesser fitness advantage. A previous study reported resistance to be

a completely dominant trait in the CAVR strain (Le Jambre et al., 2000) on the basis that no difference was observed in the resistance levels of heterozygote and homozygote resistance worms. Our finding of additivity arises from the evolution of the worm populations during passage following the backcross experiment (i.e., in BC_4.IVM.$P_{3/4}$ samples versus the BC_4.IVM data), and particularly the fixation of alleles from the resistant parent at putative segregating sites near the inferred position of resistance. In our model, alleles can reach fixation only where there is selection in favour of the homozygote resistant type over and above that for heterozygote resistant individuals.

We have here utilised a range of population genetic methods to attempt to localise selection. In addition to the standard metrics of F_{ST} and Tajima's D, we have implemented more novel methods which explicitly account for the history of the cross population. Each approach has limitations. While most models neglect linkage between adjacent sites in the genome, our multi-locus model explicitly accounts for genomic structure during the cross. However, even this model involves the use of heuristic measures to identify segregating sites between the two parental strains, and is based upon a map of recombination in the worm genome (Doyle et al., 2018) that, while the state of the art for parasitic species, was derived from a limited number of observed crossover events. None of our models accounts for genomic rearrangements during the experiment, which are inherently difficult to observe via short read sequencing.

Our ability to resolve the precise location of the variant under selection was limited by the number of recombination events in the worm population. Precise identification of the location of an allele under selection requires that the allele not be linked to other nearby alleles in at least some fraction of individuals in the population. While the inherent biological recombination rate here has a role, population bottlenecks in each round of the cross induced by the use of only a limited number of individuals - 50-100 male and female worms per cross generation - reduced our ability to resolve the region of introgression further toward a single causative gene or variant. These bottlenecks were due to the limited number of L_4 worms that it was possible to collect from each sheep and transplant in the successive generation of the cross. Genetic drift induced by successive population bottlenecks introduced considerable uncertainty in the outcome of the experiment, such that replicate sets of allele frequencies from our model showed considerable differences between each other (**Figure S5)**. This implies that the data from the experiment itself should be understood as the output of a stochastic process; beyond the clear large-scale patterns observable in the data and detailed above, more minor details of the output might not be seen again were the experiment to be repeated.

The structure of the experiment thus imposes a limit on our ability to infer the location of a selected allele, and an improved characterisation of selection would likely best be achieved through the use of alternative experimental procedures. Conducting further generations of cross as performed within this experimental framework would induce more recombination events, reducing the mean size of parental genomic blocks. However, such a course of action would be limited in scope

were it not accompanied by the use of larger L_4 populations, for example by the simultaneous passage of the population through multiple animals in parallel. Clearly, a larger or longer experiment would have cost and welfare implications: the statistical framework we have developed would help to design and justify such an approach.

Importance of chromosome-level genome assemblies for genetic mapping and population genomics

The success of identifying a single region of introgression was dependent upon a greatly improved and highly contiguous, chromosomal-scale reference genome assembly for the *H. contortus* isolate, MHco3(ISE).N1. This isolate was used in the original draft assembly of this species (Laing et al., 2013), which was derived via inbreeding of the same MHco3(ISE) strain used in the backcrosses presented. In the absence of a contiguous chromosomal scale genome assembly, interpretation of these analyses can be extremely challenging; true genetic signal(s) can be obscured by technical artifacts associated with a fragmented assembly that can include short contigs lacking spatial orientation, multiple haplotypes present, collapsed paralogs, and poor resolution of repeat structures. As a consequence, read mapping and variant calling in fragmented genomes can be suboptimal. These technical challenges are exacerbated in highly polymorphic species such as *H. contortus* and *Teladorsagia circumcincta* (Choi et al., 2017), in which reference genomes were constructed from pools of individuals that, despite inbreeding, still contained significant genetic variation. Perhaps the biggest problem with using a fragmented genome assembly for genome-wide analyses is the inability to determine linkage.

Signals of selection will often be dispersed across multiple scaffolds of a fragmented draft genome assembly when in reality they are co-located in a single genomic region. This can potentially can give rise to the erroneous conclusion that multiple independent loci show a signal of selection when it may be that only a single selected locus exists. For example, a major effect single locus that has a physically extensive signature of selection due to a hard selective sweep, and the associated hitchhiking effect, could easily be interpreted as multiple independent unlinked loci under selection. These artefacts are evident when examining the sequence data we have generated on different versions of the *H. contortus* MHco3(ISE).N1 reference genome assembly. In this case, the introgression region of chromosome V only became apparent after significant improvement in chromosome contiguity and once chromosome V was fully assembled (**Figure 7**). In the earlier, more fragmented, draft genome assemblies, the signature of selection was dispersed across multiple scaffolds making the number of major loci underlying ivermectin resistance in the MHco10(CAVR) and MHco4(WRS) strains unclear. This will be an increasingly important issue as unfinished draft reference genomes are produced for an increasing number of non-model species as whole genome sequencing becomes cheaper and more widely available. It will be important for those using genome-wide data for population genomic and genetic studies to be aware of the limitations of using unfinished draft reference genomes in various genome-wide analyses.

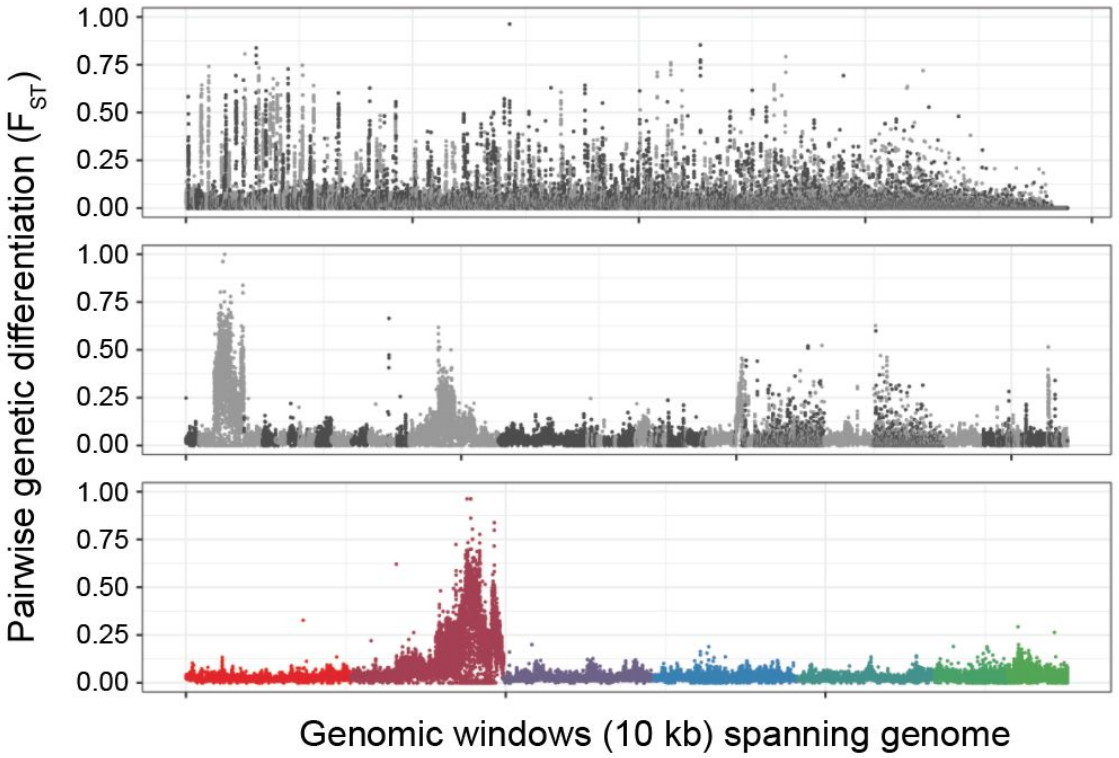


Figure 7: Demonstrable effect of genome contiguity on the resolution of the introgression region. The same pairwise comparison - MHco3(ISE) *vs* MHco3/10.BC_4.P_4 - is presented using on three versions of the *H. contortus* genome; the published V1 draft assembly (Laing et al., 2013) (top; N50 = 0.083 Mb, N50(n) = 1,151, n = 23,860), an intermediate improved assembly (middle; N50 = 5.2 Mb, N50(n) = 16, n = 6668), and the chromosomal-scale assembly presented here and elsewhere (Doyle et al., 2018)(bottom; N50 = 47.4 Mb, N50(n) = 3, n = 8). The top and middle plots are coloured light and dark grey to reflect alternate contigs / scaffolds, whereas the bottom plot is coloured as in Figure 2A.

Our comprehensive genome-wide analysis of parental strain genetic diversity re-emphasises that *H. contortus* is highly genetically diverse both within and between strains. It is clear, however, that short read mapping-based analyses that aim to exploit a single reference genome are almost certain to underestimate the

extent of genetic variation that distinguishes a divergent strain. We have exploited linked selection of a subset of this variation, i.e., only single nucleotide polymorphisms, to characterise the introgression event in these analyses; however, understanding the functional consequences of this strain specific diversity will rely on a more comprehensive description of all of the variation that defines a strain. As an alternative to a linear reference genome, the use of population genome graphs - a non-linear or branching reference that contains alternate paths representing known genetic variation - may be more suitable (Paten et al., 2017), and allow better characterisation of known variation that is too complex to be analysed using a linear reference (Maciuca et al., 2016). This could be made possible once a comprehensive analysis of population and perhaps global genetic diversity reference set of genetic variation is made available (Sallé *et al.*, in preparation). Alternatively, such variation may require geographically diverse isolates to each have a *de novo* reference sequence, ideally sequenced and assembled using long-read sequencing technology, to allow a more comprehensive description of genetic variation within a population while allowing large scale variation between reference strains to be characterised.

Conclusion

Our genome-wide analysis of two genetic crosses between ivermectin resistant and ivermectin sensitive isolates has identified a major ivermectin resistance locus in two genetically and geographically distinct strains of *Haemonchus contortus*. This result depended on the development of novel population genetic methods to understand

the genetics of the crossing procedure, and on the availability of a chromosome-scale genome assembly. More traditional population genetic analysis highlights the high level of variation both between and within the parental populations used to construct the cross, potentially explaining the difficulty in validating previous candidate genes that were identified using approaches that neither experimentally nor statistically controlled for this genetic diversity. This work represents the most comprehensive analysis of the genetics of anthelmintic resistance in a parasitic nematode to date and demonstrates the power of genetic crossing and contiguous genome assembly to eliminate false positive genetic signals typically linked to resistance. Our results eliminate the possibility that many of the previously proposed candidate genes are involved in ivermectin resistance in these isolates, and should focus future efforts on identifying the causal variant within the region we identify.

Code availability

Code used in this project is available from https://github.com/cjri/HCCross

Acknowledgement

Work at the Wellcome Sanger Institute was funded by Wellcome (grants 098051 and 206194) and by the Biotechnology and Biological Sciences Research Council (BB/M003949). CJRI was supported by a Sir Henry Dale Fellowship, jointly funded by Wellcome and the Royal Society (grant 101239/Z/13/Z). We thank Pathogen Informatics and DNA Pipelines (WSI) for their support and expertise, and Guillaume Sallé for constructive feedback on the manuscript.

Competing interests

The authors declare no competing interests.

Materials and methods

Background of strains and original backcross

The MHco3(ISE), MHco4(WRS) and MHco10(CAVR) are *H. contortus* strains maintained and stored at the Moredun Institute. The MHco3(ISE) strain was originally derived by multiple rounds of inbreeding of the SE strain (Roos et al. 2004). The precise provenance of the SE strain is not clearly recorded but is thought to be originally obtained from East Africa. MHco10(CAVR) is derived from the Chiswick Avermectin Resistant Strain (CAVR), an ivermectin resistant field strain originally isolated in Australia (Le Jambre et al., 1995). MHco4(WRS) is derived from the ivermectin resistant White River Strain (WRS) originally isolated from South Africa (van Wyk and Malan, 1988).

The experiment described here extends two previously-described backcross experiments. The construction, phenotypic validation, and initial microsatellite analysis of these backcrosses has been described previously (Redman et al., 2012), and is outlined in **Figure 1A and B.** Briefly, two independent crosses were performed in parallel: (i) MHco3(ISE) x MHco10(CAVR), and (ii) MHco3(ISE) x MHco4(WRS). In the first generation of each cross, female worms of the ivermectin resistant parental strains [MHco10(CAVR) or MHco4(WRS)] were crossed with male MHco3(ISE) worms, to generate lines designated MHco3/4 and MHco3/10, respectively (**Figure 1A**). After the first cross, the F_1 generation females were mated to male MHco3(ISE), resulting in backcross generations designated MHco3/10.BC_n and MHco3/4.BC_n, where n is the number of backcross generations (BC). This involved the recovery of immature worms from the abomasum at day 14 post

infection by necropsy of donor sheep, which were washed in physiological saline (0.85% NaCl) after which worm sex was determined (Denham, 1969), followed by surgical transfer of 45-100 male MHco3(ISE) and 50-100 female F_1 L_4/immature adult worms into worm-free recipient lambs all within 2 h of the original collection. This process of backcrossing was repeated for a total of 4 generations, i.e., MHco3/10.BC_4 and MHco3/4.BC_4 (**Figure 1B**). *In vivo* ivermectin selection was applied after mating and before collection of eggs to enrich for ivermectin resistant adults to be used in the subsequent round of the backcross. A controlled efficacy test was performed on the three adult strains and MHco3/4.BC_4 and MHco3/10.BC_4 strains to determine the initial levels of ivermectin efficacy, and the degree of resistance acquired in the introgressed lines. After four rounds of backcrossing, a selection experiment was performed. MHco3/10.BC_4 and MHco3/4.BC_4 strains were used to infect recipient sheep, after which eggs were collected and in vitro culture to L_3. These infective larvae were used to infect two sheep per strain, one that received 0.1 mg/kg ivermectin and one that remained drug naive. At 7 days post treatment, treated and naive L_4/immature adults were recovered by necropy and stored for molecular analysis.

The backcrosses described above were extended by performing a further four rounds of *in vivo* ivermectin selection, during which mating within the BC_4 population continued (**Figure 1C**). For each cross, progeny of the BC_4 generation were cultured to L_3 and used to infect a donor sheep, which was subsequently treated with ivermectin (0.1 mg/kg in round 1, followed by 0.2 mg/kg in following rounds). Eggs from ivermectin-treated survivor adults were collected post-treatment, cultured to L_3,

and used to infect a new donor sheep. L_4/immature adults were collected by necropsy after rounds three (i.e., BC_4.IVM.P_3) and four (i.e., BC_4.IVM.P_4) of selection and stored for molecular analysis.

Library preparation and sequencing

Whole genome sequencing was performed on the three parental strains (MHco3(ISE), MHco10(CAVR) & MHco4(WRS)), and from each cross, pre and post ivermectin treatment following four rounds of backcrossing (MHco3/10.BC_4.noIVM, MHco3/4.BC_4.noIVM, MHco3/10.BC_4.IVM, MHco3/4.BC_4.IVM), and after 3 and 4 rounds of additional selection after the backcross (MHco3/10.BC_4.IVM.P_3, MHco3/4.BC_4.IVM.P_3, MHco3/10.BC_4.IVM.P_4, MHco3/4.BC_4.IVM.P_4). Pools of male and female worms were included for the parental (n = 50-60 worms) and BC_4 samples (n = 25-30 worms), whereas only female worms were used in the passage 3 & 4 samples (n = 60 worms). Genomic DNA from each pooled sample was prepared using a phenol chloroform extraction protocol as previously described. Sequencing libraries were prepared using a PCR-free protocol (Kozarewa et al. 2009), and sequenced as described in **Table S1**. We generated approximately 6.14×10^{11} bp of sequence data, which equates to approximately 199.65× unmapped genome coverage per sample. Raw sequence data quality was analysed using FASTQC (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)* and was assessed using MultiQC (Ewels et al., 2016) prior to further processing.

The raw sequencing data generated during the current study are available in the European Nucleotide Archive repository (http://www.ebi.ac.uk/ena/) under the study accession number PRJEB2353 (**Table S1**).

Mapping and variant analysis

Raw sequence data was mapped per lane to the current reference genome (available here: ftp://ngs.sanger.ac.uk/production/pathogens/Haemonchus_contortus) using BWA-MEM (Li, 2013)(default parameters with *-k 15*). Samples for which multiple lanes of mapped data were available were merged, duplicate reads were marked and removed using *Picard v2.5.0* (https://github.com/broadinstitute/picard), and mapped perfect read pairs were extracted per sample (*samtools-1.3 view -f 2*).

Genome-wide variants were determined for each sample using samtools-1.3 mpileup (*-F 0.25 -d 500*). Nucleotide diversity and Tajima's D was determined using *npstat* (Ferretti et al., 2013), which required pileup files (which we derived from the mpileup generated above) that were split per sample per chromosome as input. We determined short-range linkage disequilibrium between pairs of variants in single and paired reads using *LDx* (Feder et al., 2012). We compared these data with estimates of LD decay over genetic distance as proposed by Hill and Weir (Feder et al., 2012; Hill and Weir, 1988). *Popoolation2* (Kofler et al., 2011) was used for the analysis of pairwise genetic diversity throughout the genome. Briefly, the mpileup file was converted into a synchronised file (*popoolation2 mpileup2sync.jar --min-qual 20*), after which indels + 5 bp (popoolation2 *identify-indel-regions.pl −−min-count 2 −−indel-window 5*) and repetitive and difficult to map regions (separately identified by

repeat masking the genome [http://www.repeatmasker.org/]) were excluded (*popoolation2 filter-sync-by-gtf.pl*). The synchronised file was used as input to determine pairwise F_{ST} which was calculated in 10 kb windows throughout the genome (*popoolation2 fst-sliding.pl --pool-size 50 --window-size 10000 --step-size 5000 --min-count 4 --min-coverage 50 --max-coverage 2%*). Similarly, per base comparisons were determined using a Fisher's exact test (*popoolation2 fisher-test.pl --min-count 4 --min-coverage 50 --max-coverage 2% --suppress-noninformative*).

Copy number variation between parental strains was determined using *cnv-seq* (Xie and Tammi, 2009). Best read mapping hits were determined per bam (*samtools-1.3 view -F 4 bam | perl -lane 'print "$F[2]\t$F[3]"' > bam.hits*), before read counts for pairwise comparisons determined in 10 kb sliding windows using *cnv-seq.pl*. The R package *cnv* (v 0.2-8) was used to determine pairwise chromosome-wide normalised $\log_2$ ratios. Characterisation of structural variation in the parental strains was performed using the *speedseq sv* pipeline (Chiang et al., 2015) to identify putative duplications, deletions and inversions in each strain. This approach exploits all reads mapped using BWA-MEM, including split and discordant read pairs, which are subsequently scored using LUMPY and SVTyper. Conservative filtering was applied to retain only homozygous variants (1/1) with a minimum quality score of 100.

Population genetic modelling

Bespoke single-locus and multi-locus models were used for population genetic inference from the data. Common to both methods, variants in the genome potentially associated with resistance were identified. At each locus in the genome, a

nucleotide {A,C,G,T} was defined as existing in the susceptible parental population if it was observed at a frequency of at least 1% in that population. Loci in the resistant parental population were then identified for which exactly one nucleotide that did not exist in the susceptible population was observed at a frequency of at least 1%; without prejudice as to its phenotypic effect, this was denoted the 'resistant' allele at that locus, being associated with the resistant parental population.

Single-locus model

A single-locus population genetic model was then used to identify variants with frequencies that were inconsistent with selective neutrality. The neutral expectation was calculated using a Wright-Fisher evolutionary model to simulate the progress of a variant allele, modelling the probability distribution of the frequencies of homozygote resistant (q_1), heterozygote (q_h), and homozygote susceptible (q_0) individuals throughout the course of the experiment under the assumption of selective neutrality. The Wright-Fisher model uses a simple multinomial process for propagation; if the next generation of the population contains N individuals, the probability of obtaining n_1 , n_h , and n_0 individuals with each diploid variant is given by:

$$P(n_1, n_h, n_0) = \frac{N!}{n_1! n_h! n_0!} q_1^{n_1} q_h^{n_h} q_0^{n_0} \qquad [1]$$

At the stages in the experiment where the population was comprised of eggs, we assumed N to be large, such that the processes of crossing and backcrossing could

be described deterministically. Under these circumstances the genotype frequencies following a cross, denoted q_1', q_h', and q_0', are given by:

$$
\begin{aligned}
q_1' &= q_1^2 + q_1 q_h + \tfrac{1}{4} q_h^2 \\
q_h' &= q_1 q_h + 2 q_1 q_0 + \tfrac{1}{2} q_h^2 + q_h q_0 \\
q_0' &= q_0^2 + \tfrac{1}{4} q_h^2
\end{aligned}
\qquad [2]
$$

while the genotype frequencies following a backcross with the homozygous susceptible strain are given by:

$$
\begin{aligned}
q_1' &= 0 \\
q_h' &= q_1 + \tfrac{1}{2} q_h \\
q_0' &= q_0 + \tfrac{1}{2} q_h
\end{aligned}
\qquad [3]
$$

Evaluating this process gives a probability distribution for the frequency of the resistant allele at any given point in the experiment, dependent upon the number of resistant alleles in the initial resistant population. This number of resistant alleles, which we denote n_r, is unknown; sequence data from the resistant parental strain were used to generate a prior distribution for this value.

We first suppose that the frequency of the resistant allele in the resistant parental population is equal to some value, p_r. Given that the initial experimental population

contains 50 resistant worms, with 100 alleles, collected from this larger population, the distribution of n_r is then given by:

$$P(n_r = j) = \binom{100}{j} p_r^j (1 - p_r)^{100-j} \qquad [4]$$

We now suppose that sequencing the resistant parental population gave a_r resistant alleles at the locus in question out of a total read depth of A_r. The initial frequency p_r is unknown; however using the data an estimate can be made for this statistic, expressed as a distribution of the allele frequency. Specifically, under the assumption of a uniform prior, the underlying probability p_r can be said to be distributed as a beta distribution with parameters a_r and $A_r - a_r + 1$. We therefore obtain the following distribution for the statistic n_r:

$$P(n_r = j) = \int_0^1 \frac{A_r!}{n!(A_r - a_r!)} p_r^{a_r+1} (1 - p_r)^{A_r - a_r + 1} \binom{100}{j} p_r^j (1 - p_r)^{100-j} dp_r \qquad [5]$$

for any value of j between 0 and 100. Values of this distribution were calculated using numerical integration. This process generated the neutral expectation of the resistant allele frequency conditional on the observation of this frequency in the resistant population. If at the sampling point t, a total of N_t worms were collected, the distribution of the number of worms n_t with the resistant allele at that point is given by:

$$P(n_t = k) = \sum_j P(n_t | n_r = j) P(n_r = j) \quad [6]$$

for all $0 \leq k \leq N_t$.

The extent to which observed allele frequencies were consistent with the neutral model was calculated using a likelihood model:

$$L_t = \sum_{k=0}^{N_t} P(n_r = k) \binom{A_t}{a_t} (k/N_t)^{a_t} (1 - (k/N_t))^{A_t - a_t} \quad [7]$$

where at time t, a_t resistant alleles from a read depth of A_t were observed. A low likelihood L_t indicates deviation from the neutral expectation; Bonferroni correction was used to identify significance at the 95% level. Data from each sample were analysed; four-fold non-neutral sites, being loci for which a significant likelihood was calculated from all four samples collected throughout each cross experiment, were identified.

Multi-locus model

A multi-locus model was developed to describe the manner in which allele frequencies would be expected to change over the course of the experiment; this model exploits information about the location of variant loci provided in the *H. contortus* reference genome, and the genome-wide map of recombination rate in the worms (Doyle et al., 2018); this map was inferred by characterising recombination breakpoints in F_1 L_3 progeny by whole genome sequencing, from which

recombination rates were determined by comparing genetic distances between SNPs against their physical location in the genome. Where fixed genetic differences occur between parental strains, the dynamics of adaptation in the resulting cross are relatively straightforward (Illingworth and Mustonen, 2013; Illingworth et al., 2012); this has been exploited to identify quantitative trait loci in yeast and malaria cross populations (Abkallo et al., 2017; Parts et al., 2011). Here a heuristic approach was used to identify fixed genetic differences between the parental strains before modelling evolution at these loci to identify the location of alleles conveying drug resistance.

Three filters were used to identify putatively fixed differences between parental strains. Firstly, we identified genomic sites for which the frequency of the resistant allele was 95% or greater in the resistant parent, requiring a read depth of at least 50x coverage for such sites to achieve a good level of statistical certainty. Secondly, we noted that, following the backcross performed at the start of the experiment, no locus can be homozygous for the resistant allele and as such, the resistant allele frequency can be no greater than 50% in the population; by way of reducing noise, we removed any locus with resistant alleles at a frequency of 60% or greater in the MHco3/10.BC_4.noIVM sequence data. Thirdly, we noted that at homozygous separating sites in a population following a cross, the allele frequency will change smoothly over time; where there are N genomes in the population, the allele frequency, considered as a function of chromosome position, will change by $1/N$ wherever a crossover recombination event occurs within a genome. A diffusion model of allele frequency change, described in a previous publication (Abkallo et al.,

2017), was used to identify allele frequencies across the genome that were consistent with this pattern. This analysis fits a posterior distribution to the allele frequencies across the genome, also inferring the extent of noise in the sequence data via a beta-binomial model intrinsic to the fitting process. Conservatively, sites no more than eight standard deviations from the mean of the posterior frequencies were included in the analysis. These approaches resulted in data from a total of 1,368 loci in the MHco10(CAVR) dataset, and from 219 loci in the MHco4(WRS) dataset, were retained; the full set of allele frequencies are shown in Supporting Figure S5.

Considering these data, we used an individual-based Wright-Fisher simulation to model the outcome of the experiment, taking into account the backcrossing, selection, and bottlenecking according to the experimental design. This approach accounted for the stochastic nature of allele frequency changes due to the effect of repeated population bottlenecking. We suppose that selection acts in favour of the resistance allele at a given locus when worms are in a sheep being treated with ivermectin, such that the fitnesses of individuals that are homozygous for the resistant allele (w_1), heterozygous for the resistant allele (w_h), or homozygous for susceptible alleles (w_0) are given by:

$$\begin{aligned} w_1 &= 1 + s \\ w_h &= 1 + hs \\ w_0 &= 1 \end{aligned} \qquad [8]$$

where s is the selection coefficient (fitness advantage of the homozygous resistant compared to homozygous susceptible) and h is the dominance coefficient (proportion of fitness change conveyed by a single copy of the allele). We assumed that all worms have equal fitness in the absence of the drug. The likelihood of any given inference was calculated using the beta-binomial likelihood function obtained from the diffusion model above. Accounting for the stochasticity of the system, a set of at least 250 simulations were generated for each set of parameters, calculating the mean likelihood fit to the data across the simulations. A bootstrapping process was applied to quantify the uncertainty in this likelihood.

In an extension to this model, a two-driver was considered, in which the resistant variant at each of two loci was under selection; to allow the exploration of model space in this case within reasonable computational time the assumption was made of additive selection at each locus. To distinguish this from the previous model, a requirement was imposed that the selected alleles be separated by at least 2 Mb in the genome.

Table S1. Sample sequencing data archived at European Nucleotide Archive repository under the study accession PRJEB2353

- See accompanying xlsx document

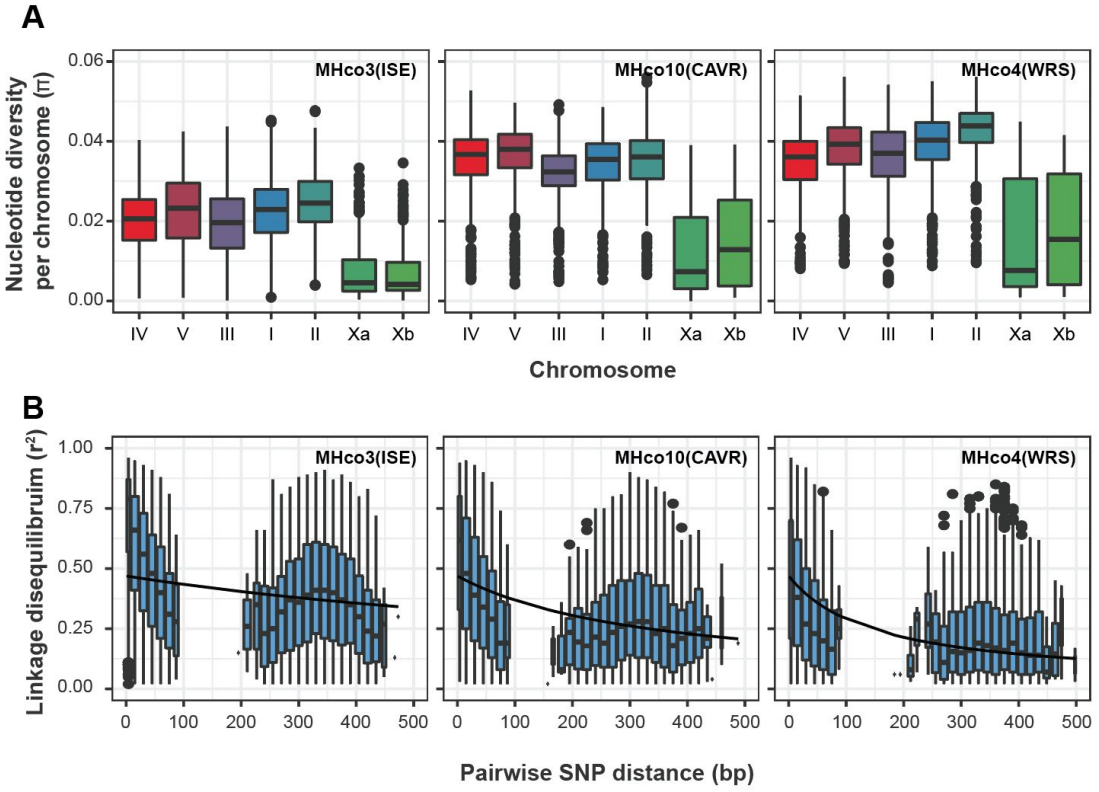


Figure S1. Characterisation of within-strain diversity. A. Within strain nucleotide diversity per chromosome, summarising genome-wide data presented in Figure 2C. Colours represent chromosomes as described in Figure 2A. **B**. Linkage disequilibrium between variants present in paired reads was estimated using LDx for each parental strain. Line represents expected LD decay over genetic distance (Feder et al., 2012; Hill and Weir, 1988).

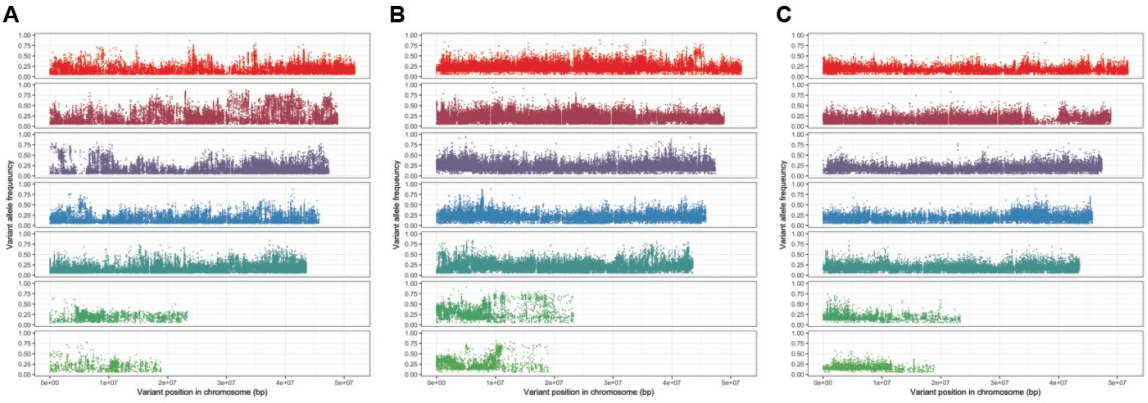
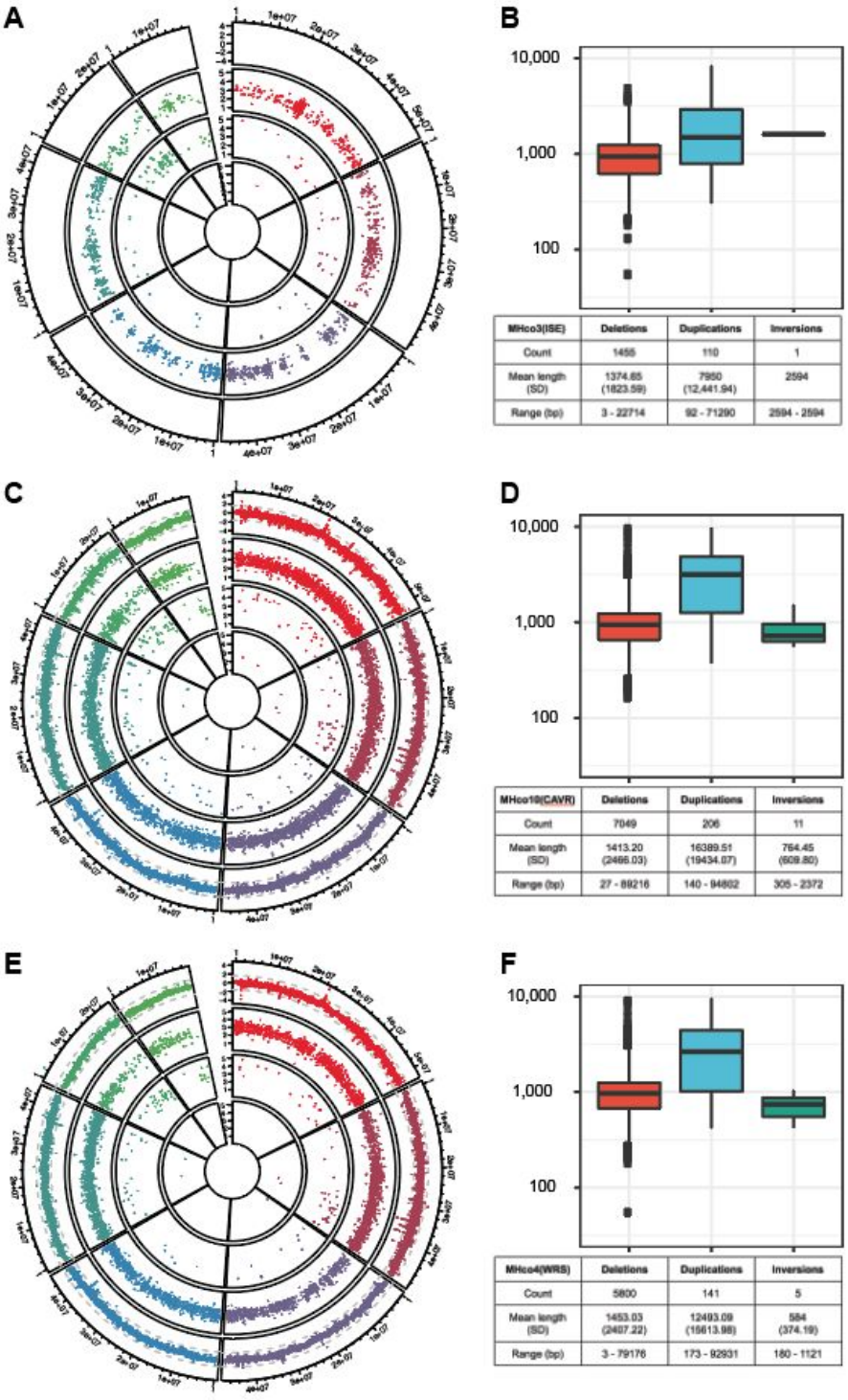


Figure S2. Distribution of "private" variant sites per parental population. A. MHco3(ISE). **B**. MHco10(CAVR). **C**. MHco4(WRS). Private sites were defined as having a frequency greater than 0.05 in the population of interest, but less than 0.05 in the two additional populations.



MHco3(ISE)	Deletions	Duplications	Inversions
Count	1455	110	1
Mean length (SD)	1374.65 (1823.59)	7950 (12,441.94)	2594
Range (bp)	3 - 22714	92 - 71290	2594 - 2594

MHco10(CAVR)	Deletions	Duplications	Inversions
Count	7049	206	11
Mean length (SD)	1413.20 (2466.03)	16389.51 (19434.07)	764.45 (609.80)
Range (bp)	27 - 89216	140 - 94802	305 - 2372

MHco4(WRS)	Deletions	Duplications	Inversions
Count	5800	141	5
Mean length (SD)	1453.03 (2407.22)	12493.09 (15613.98)	584 (374.19)
Range (bp)	3 - 79176	173 - 92931	180 - 1121

Figure S3. Copy number and structural variation in the parental lines. A,B. MHco3(ISE), **C,D**. MHco10(CAVR), **E,F.** MHco4(WRS). Data per circos plot (**A,C,E**) is orientated as follows; outer circle: CNV variation between MHco3(ISE) and each resistant parent. No CNV comparison was made in A; second circle: deletions; third circle: duplications; inner circle: inversions. The data presented in **A,C,E** is summarised in the boxplots (feature length distribution) and tables in **B,D,F**.

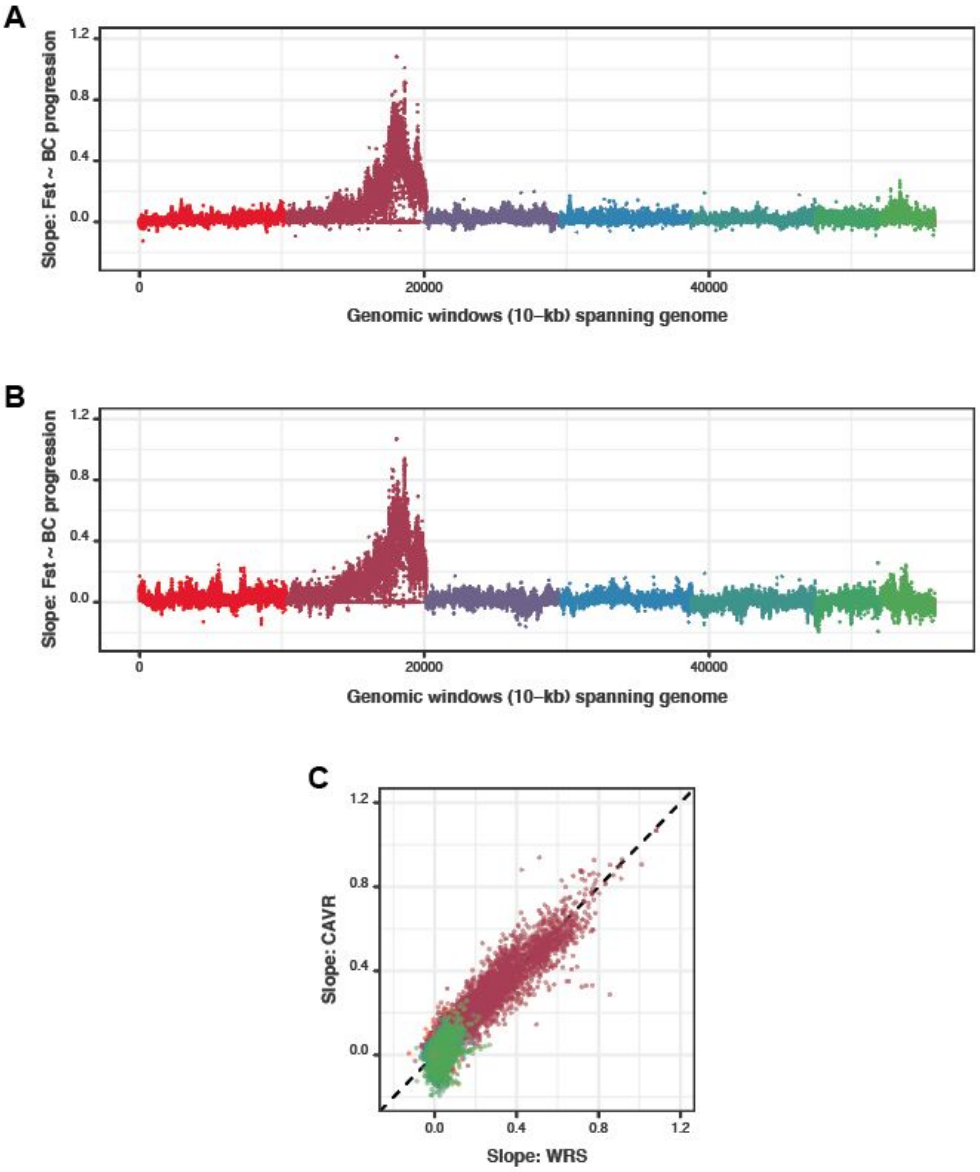


Figure S4. Summary of genome-wide change in F_{ST} throughout the backcross and subsequent passage. Linear regression between F_{ST} and the four sampling time points was performed for each 10 kb window sampled across the genome for both MHco3/10 (**A**) and MHco3/4 (**B**). The slope of the regression was plotted. Panel **C** shows the correlation between the slopes (F_{ST} vs backcross progression) for MHco3/10 (A) and MHco3/4 (B). The dashed line represents x=y. Colours represent chromosomes as described in Figure 2A.

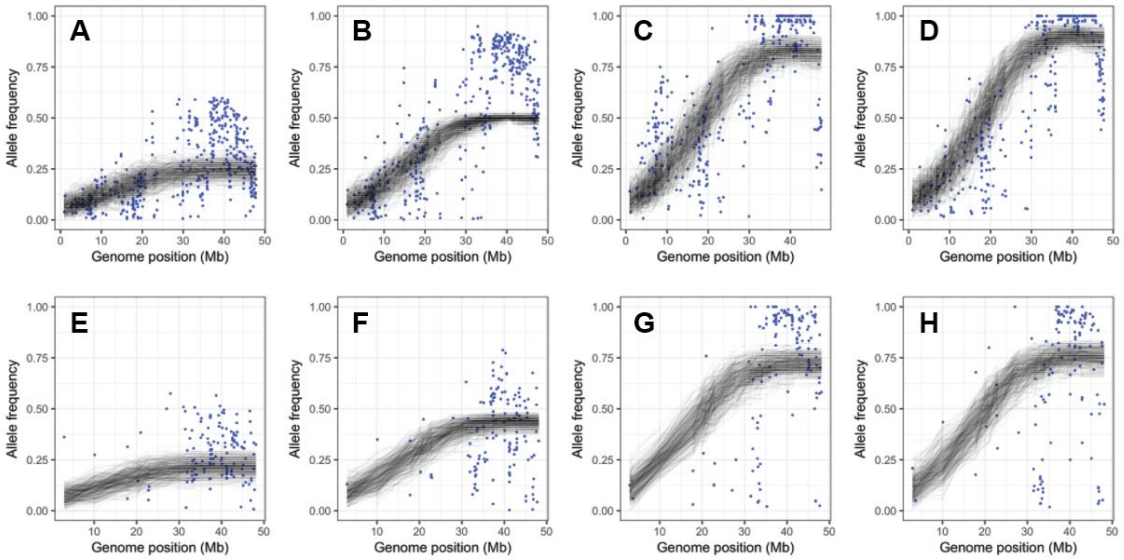


Figure S5. Fits between the model and the data for each data sample. Blue dots show filtered allele frequencies for putative segregating sites. The model fit is shown as gray lines; a distinct line is shown for each of the 250 replicate simulations run for the parameters generating the maximum likelihood fit. **A.** MHco3/10.BC_4.noIVM. **B**. MHco3/10.BC_4.IVM. **C.** MHco3/10.BC_4.IVM.P_3. **D.** MHco3/10.BC_4.IVM.P_4. **E.** MHco3/4.BC_4.noIVM. **F**. MHco3/4.BC_4.IVM. **G.** MHco3/4.BC_4.IVM.P_3. **H.** MHco3/4.BC_4.IVM.P_4.

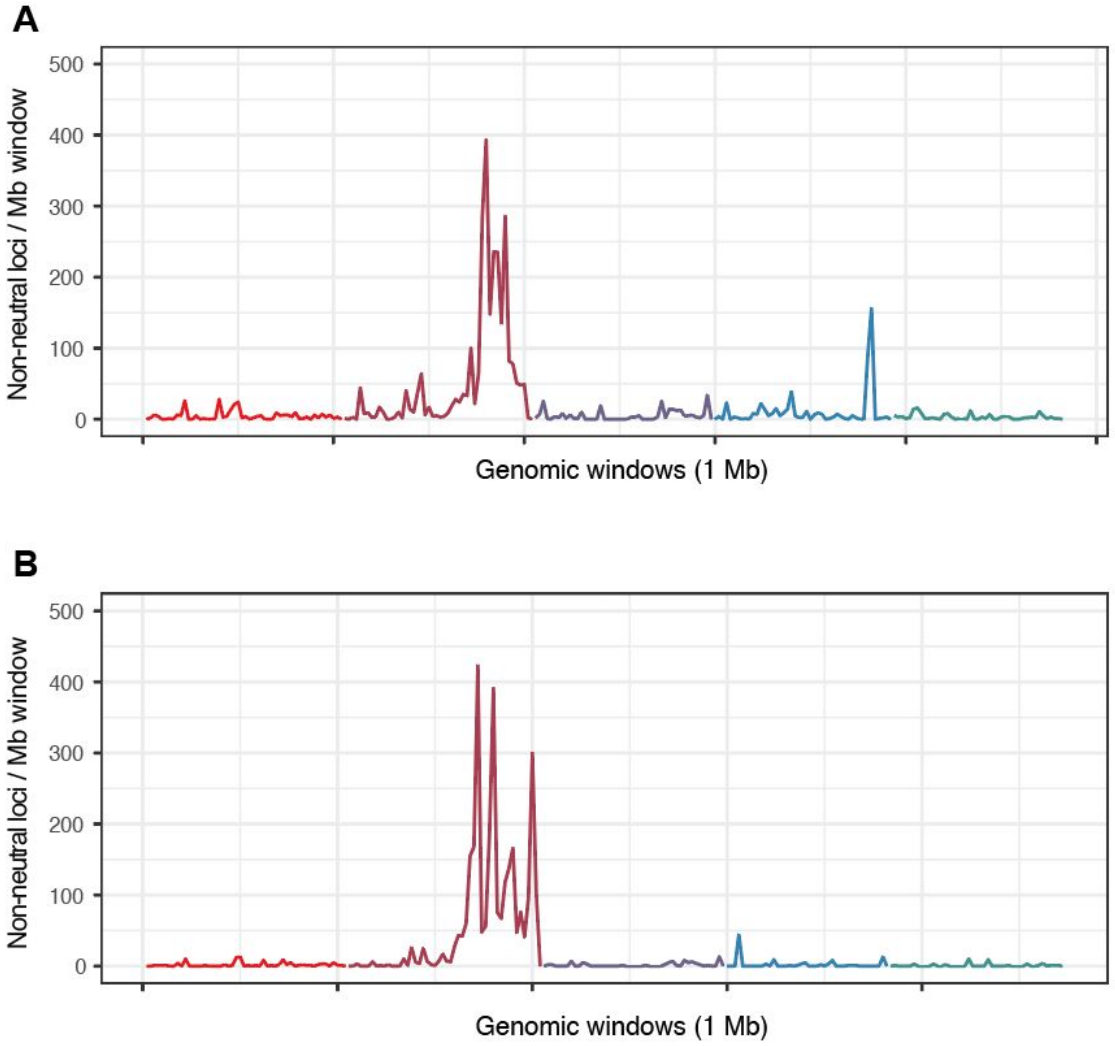


Figure S6. Location of significantly non-neutral loci identified using the single-locus population genetic model. Corresponding peaks in the location of significant sites can be seen in the MHco3/10 (**A**) and MHco3/4 datasets (**B**). A total of 70.6% of significant sites in the MHco3/10 dataset, and 90.6% of significant sites in the MHco3/4 dataset, were found in chromosome V. Data are binned in 1 Mb windows spanning the genome. Colours represent chromosomes as described in Figure 2A.

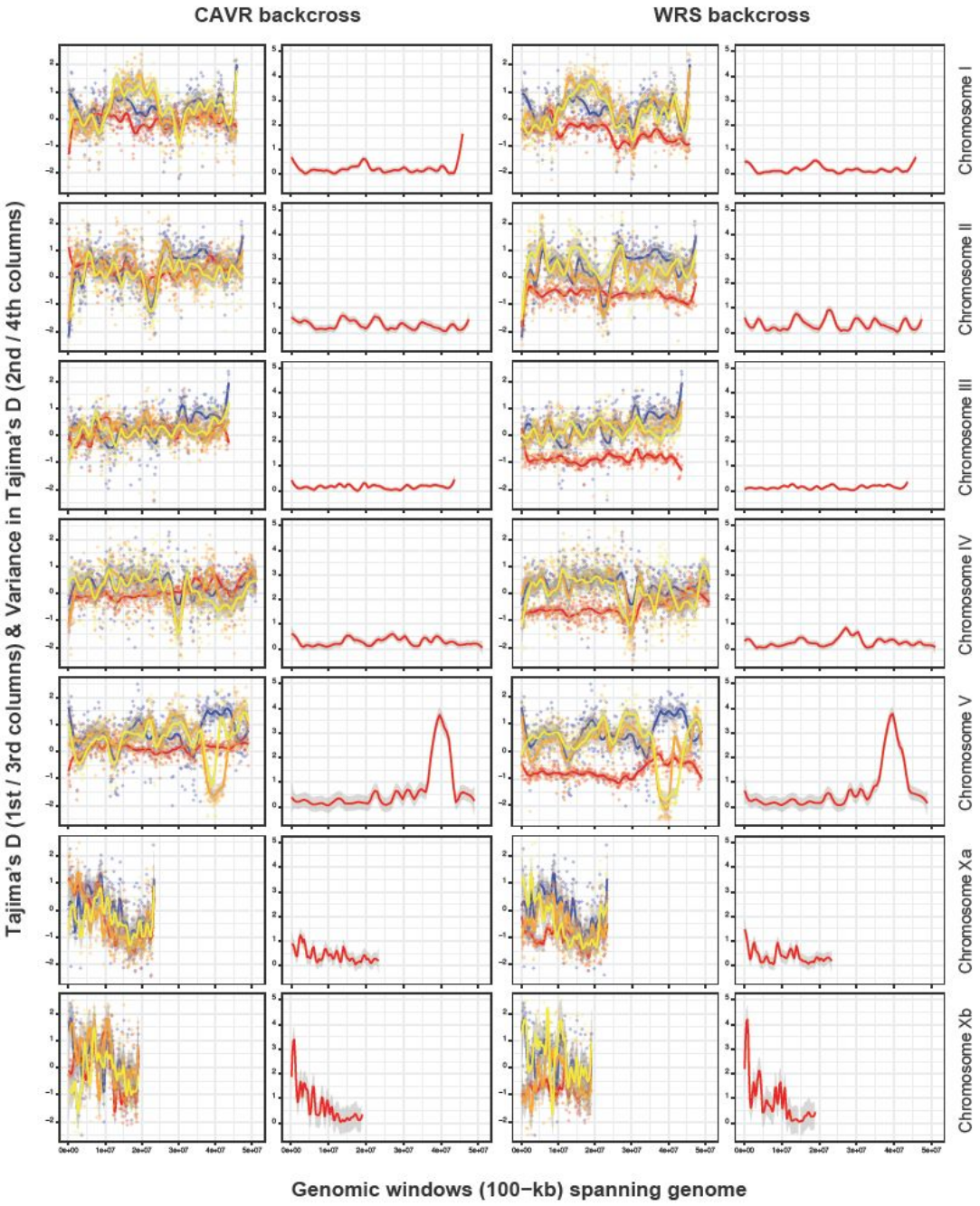


Figure S7. Analysis of Tajima's D variation in each chromosome per cross. Comparison of Tajima's D per chromosome between MHco3(ISE) parent (blue), MHco10(CAVR) (panel column 1; red) or MHco4(WRS) (panel column 3; red) and passages 3 (orange) and 4 (yellow) of the crosses. Tajima's D was calculated using *npstat* in 100 kb windows spanning the genome. The variance in the mean value of Tajima's D among MHco3(ISE) and passages 3 and 4 – for which an increase in variance would suggest introgression and evidence of selection – was determined and is presented as smoothed line (red) in panel columns 2 and 4.

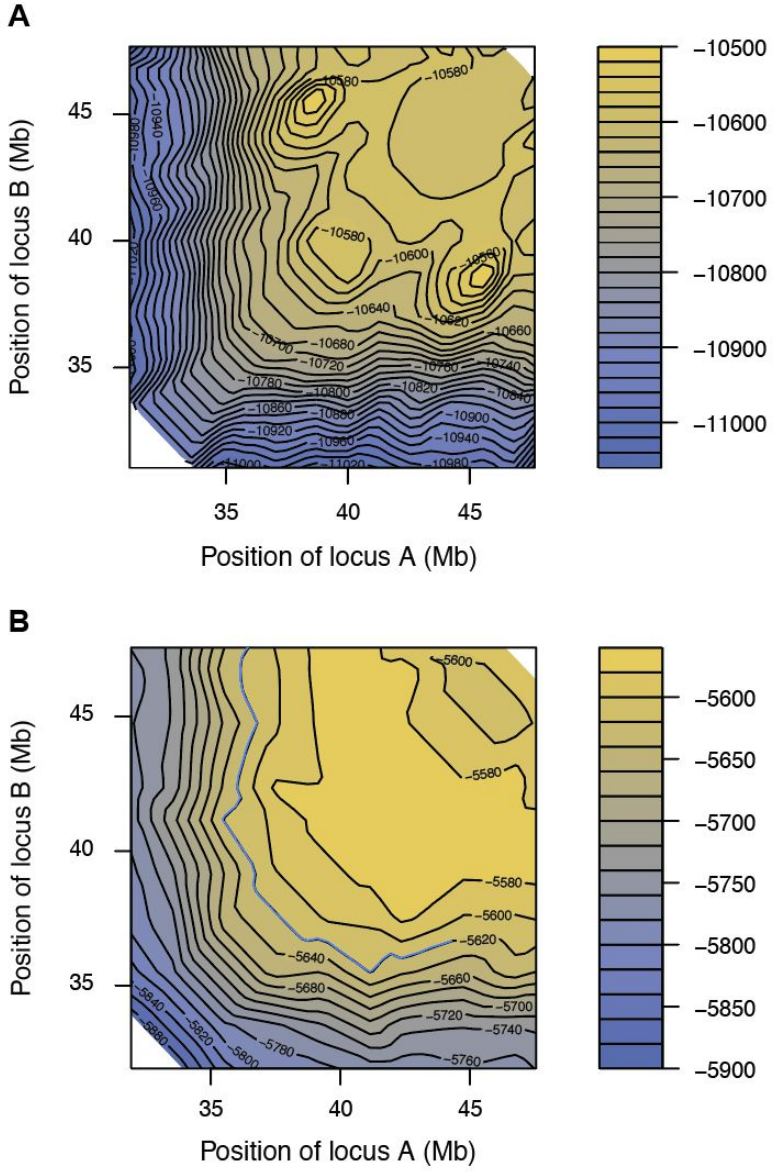


Figure S8. Contour maps of log likelihood scores derived from the two locus driver model. **A.** MHco10(CAVR). **B.** MHco4(WRS). The model was restricted to interactions between pairs of loci at least 2 Mb apart.

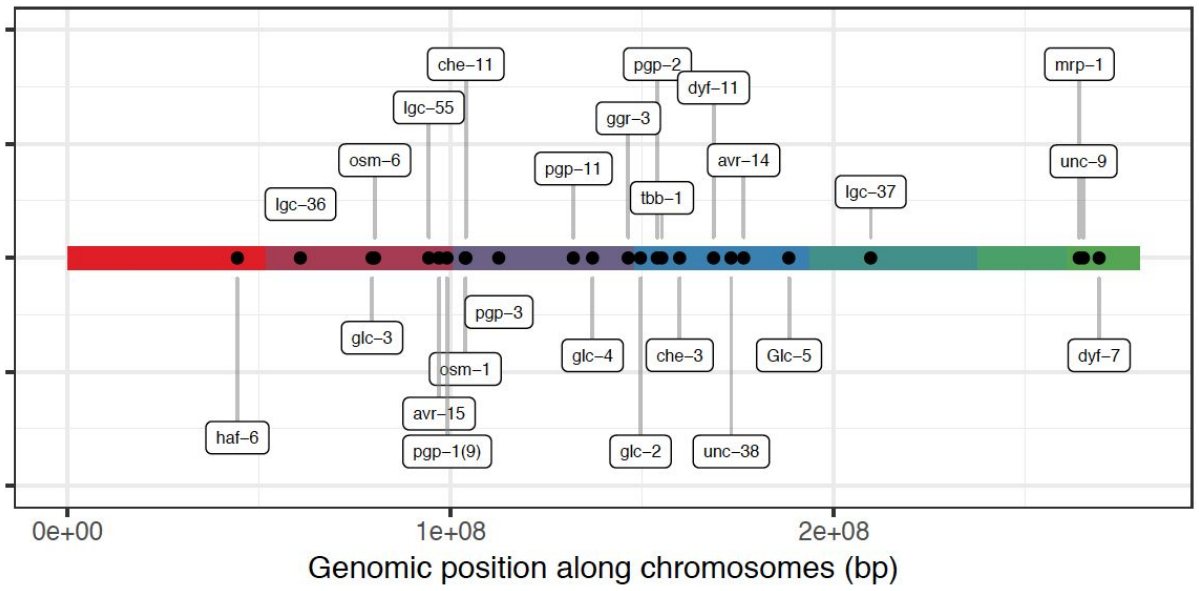


Figure S9. Relative position and location of candidate genes from the literature proposed to be associated with ivermectin resistance in *Haemonchus contortus* and / or *Caenorhabditis elegans.* Gene coordinates are presented in Table S2. Colours represent chromosomes as described in Figure 2A.

Table S2. Candidate genes from literature proposed to be associated with ivermectin resistance in *Haemonchus contortus* and / or *Caenorhabditis elegans*

Gene	GeneID [3]	Chromosome	Coordinates [4]	Reference
avr-14b / gbr-2 or HcGluClα3 [2]	HCOI00274000, HCOI00274100	I	28352540 - 28362260	(Williamson et al., 2011)
avr-15	HCOI00130400	V	45145104 - 45158307	(Dent et al., 2000)
tbb-1	HCOI01967600	I	7026384 - 7029311	(Eng et al., 2006; de Lourdes Mottier and Prichard, 2008)
che-3 [1]	HCOI01710300	I	11666953 - 11695321	(Dent et al., 2000)
dyf-11 [1]	HCOI01514900	I	20494101 - 20505946	(Dent et al., 2000)
dyf-7	HCOI01065900	Xb	8461163 - 8465904	(Urdaneta-Marquez et al., 2014)
ggr-3	HCOI00098800	II	45626226 - 45639348	(Rao et al., 2009)
glc-1 [1]	NA			(Dent et al., 2000)
glc-2 / GluClb [2]	HCOI00383000	I	1443439 - 1446121	
glc-3	HCOI00543300	V	27643249 - 27655188	(Williamson et al., 2011)
glc-4	HPLM_0001206301	II	36314509 - 36331683	
glc-5 / HcGluCla [2]	HCOI00617300	I	40196936 - 40206359	(Blackhall et al., 1998)
lgc-36	HCOI02054500	V	9008092 - 9028181	
lgc-37 / HG-1 [2]	HCOI00977100	III	15741515 - 15762899	(Blackhall et al., 2003)
lgc-55	HCOI00162900	V	42450788 - 42469177	
osm-1 [1]	HCOI00970500	II	3105809 - 3126440	

osm-3	HCOI00164700, HCOI00673700			
osm-5 [1]	HCOI00498400, HCOI00077900			
pgp-1 / pgp-9 [2]	HCOI00233200	V	47249715 - 47265234	(James and Davey, 2009; Raza et al., 2016; van Wyk and Malan, 1988)
pgp-2 / pgp-a [2]	HCOI00025600	I	5849505 - 5868032	(Raza et al., 2016)
pgp-11	HCOI00622400	II	31321675 - 31344154	(Raza et al., 2016)
unc-38	HCOI01939400	I	25118411 - 25125593	
unc-7 [1]	HPLM_0000412001			(Dent et al., 2000)
unc-9 [1]	HCOI02075100	Xb	4410309 - 4424525	(Dent et al., 2000)
che-11	HCOI00062600	II	3237676 - 3245201	
osm-6	HCOI00937900	V	28370676 - 28397433	
pgp-3	HCOI00117000	II	11888372 - 11913805	(Raza et al., 2016)
haf-6	HCOI01007100	IV	44336050 - 44341372	(Raza et al., 2016)
pgp-13 / pgp-12 [1]	HCOI01115500	II	10272077 - 10286519	(David et al., 2018)
mrp-1 [1]	HCOI01904000	Xb	3325452 - 3334565	(James and Davey, 2009)

1. *C. elegans* genes
2. Alternate names in the literature
3. Gene IDs correspond to the annotated gene in the published version of the MHco3(ISE) genome (Laing et al., 2013).
4. Coordinates are approximate, as they represent the transfer of coordinates from the published version based on homology, and therefore may have changed depending on the conservation of the gene structure in the genome between the published and updated genome version used here.

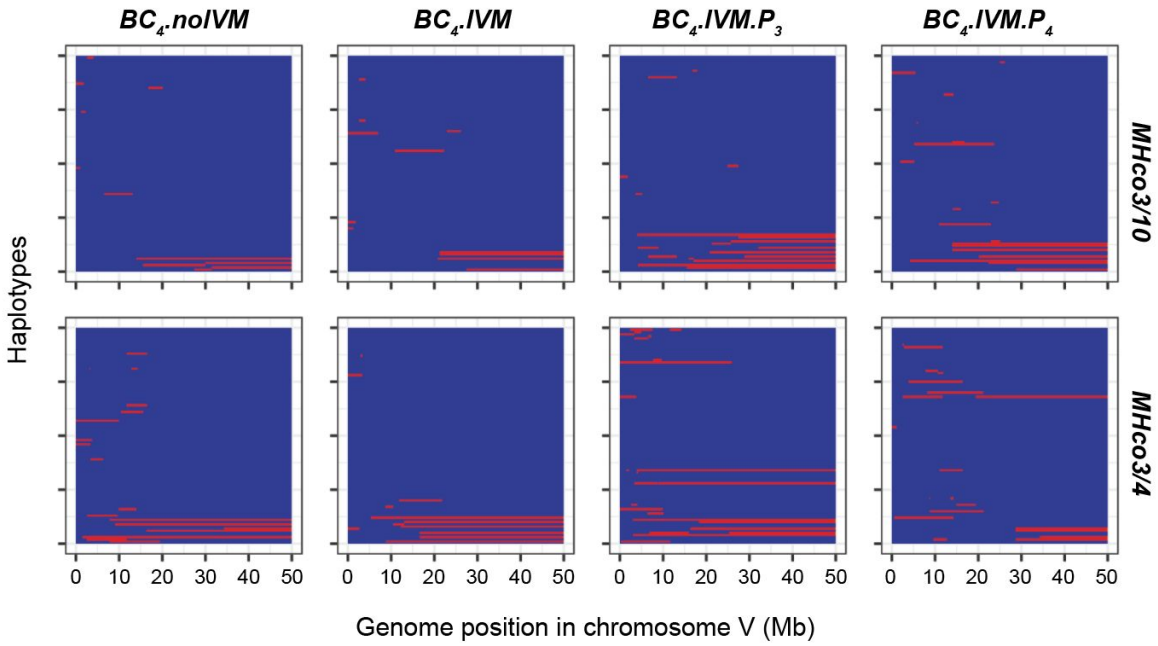


Figure S10: Haplotype structure of chromosome V in an example output from the model under neutral evolution. Segments of genome from the resistant parent are shown in red, while segments of genome from the susceptible parent are shown in blue. The repeated backcross removes most of the resistant genotypes from the population.

References

Abkallo, H.M., Martinelli, A., Inoue, M., Ramaprasad, A., Xangsayarath, P., Gitaka, J., Tang, J., Yahata, K., Zoungrana, A., Mitaka, H., et al. (2017). Rapid identification of genes controlling virulence and immunity in malaria parasites. PLoS Pathog. *13*, e1006447.

Blackhall, W.J., Pouliot, J.-F., Prichard, R.K., and 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., and Beech, R.N. (2003). Selection at a γ-aminobutyric acid receptor gene in Haemonchus contortus resistant to avermectins/milbemycins. Mol. Biochem. Parasitol. *131*, 137–145.

Britton, C., Roberts, B., and Marks, N.D. (2016). Functional Genomics Tools for Haemonchus contortus and Lessons From Other Helminths. Adv. Parasitol. *93*, 599–623.

Chiang, C., Layer, R.M., Faust, G.G., Lindberg, M.R., Rose, D.B., Garrison, E.P., Marth, G.T., Quinlan, A.R., and Hall, I.M. (2015). SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat. Methods *12*, 966–968.

Choi, Y.-J., Bisset, S.A., Doyle, S.R., Hallsworth-Pepin, K., Martin, J., Grant, W.N., and Mitreva, M. (2017). Genomic introgression mapping of field-derived multiple-anthelmintic resistance in Teladorsagia circumcincta. PLoS Genet. *13*, e1006857.

Crellen, T., Walker, M., Lamberton, P.H.L., Kabatereine, N.B., Tukahebwa, E.M., Cotton, J.A., and Webster, J.P. (2016). Reduced Efficacy of Praziquantel Against Schistosoma mansoni Is Associated With Multiple Rounds of Mass Drug Administration. Clin. Infect. Dis. *63*, 1151–1159.

David, M., Lebrun, C., Duguet, T., Talmont, F., Beech, R., Orlowski, S., André, F., Prichard, R.K., and Lespine, A. (2018). Structural model, functional modulation by ivermectin and tissue localization of Haemonchus contortus P-glycoprotein-13. Int. J. Parasitol. Drugs Drug Resist.

Denham, D.A. (1969). The Development of Ostertagia circumcincta in Lambs. J. Helminthol. *43*, 299.

Dent, J.A., Smith, M.M., Vassilatis, D.K., and Avery, L. (2000). The genetics of ivermectin resistance in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *97*, 2674–2679.

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., et al. (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.D., Holroyd, N., Mable, B.K., Maitland, K., Morrison, A.A., et al. (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.

Echevarria, F., and Trindade, G. (1989). Anthelmintic resistance by Haemonchus contortus to ivermectin in Brazil: a preliminary report. Vet. Rec. *124*, 147–148.

Elard, L., and Humbert, J.F. (1999). Importance of the mutation of amino acid 200 of the isotype 1 β-tubulin gene in the benzimidazole resistance of the small-ruminant parasite Teladorsagia circumcincta. Parasitol. Res. *85*, 452–456.

Eng, J.K.L., Blackhall, W.J., Osei-Atweneboana, M.Y., Bourguinat, C., Galazzo, D., Beech, R.N., Unnasch, T.R., Awadzi, K., Lubega, G.W., and Prichard, R.K. (2006). Ivermectin selection on β-tubulin: Evidence in Onchocerca volvulus and Haemonchus contortus. Mol. Biochem. Parasitol. *150*, 229–235.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

Feder, A.F., Petrov, D.A., and Bergland, A.O. (2012). LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS One *7*, e48588.

Ferretti, L., Ramos-Onsins, S.E., and Pérez-Enciso, M. (2013). Population genomics from pool sequencing. Mol. Ecol. *22*, 5561–5576.

Geary, T.G. (2016). Haemonchus contortus: Applications in Drug Discovery. Adv. Parasitol. *93*, 429–463.

Geerts, S., and Gryseels, B. (2000). Drug Resistance in Human Helminths: Current Situation and Lessons from Livestock. Clin. Microbiol. Rev. *13*, 207–222.

Gilleard, J.S. (2013). Haemonchus contortus as a paradigm and model to study anthelmintic drug resistance. Parasitology *140*, 1506–1522.

Hefnawy, A., Berg, M., Dujardin, J.-C., and De Muylder, G. (2017). Exploiting Knowledge on Leishmania Drug Resistance to Support the Quest for New Drugs. Trends Parasitol. *33*, 162–174.

Hill, W.G., and Weir, B.S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. *33*, 54–78.

Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite − a comprehensive resource for helminth genomics. Mol. Biochem. Parasitol. *215*, 2–10.

Hunt, P. (1995). Identification and characterisation of dominant ivermectin resistant genes in Caenorhabditis elegans. PhD. University of New England.

Illingworth, C.J.R., and Mustonen, V. (2013). Quantifying selection in evolving populations using time-resolved genetic data. J. Stat. Mech: Theory Exp. *2013*, P01004.

Illingworth, C.J.R., Parts, L., Schiffels, S., Liti, G., and Mustonen, V. (2012). Quantifying selection acting on a complex trait using allele frequency time series data. Mol. Biol. Evol. *29*, 1187–1197.

Jabbar, A., Campbell, A.J.D., Charles, J.A., and Gasser, R.B. (2013). First report of anthelmintic resistance in Haemonchus contortus in alpacas in Australia. Parasit. Vectors *6*, 243.

James, C.E., and 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.

Kaplan, R.M., and Vidyashankar, A.N. (2012). An inconvenient truth: global worming and anthelmintic resistance. Vet. Parasitol. *186*, 70–78.

Kofler, R., Pandey, R.V., and Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics *27*, 3435–3436.

Kotze, A.C., Hunt, P.W., Skuce, P., von Samson-Himmelstjerna, G., Martin, R.J., Sager, H., Krücken, J., Hodgkinson, J., Lespine, A., Jex, A.R., et al. (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.

Kwa, M.S.G., Veenstra, J.G., and Roos, M.H. (1994). Benzimidazole resistance in Haemonchus contortus is correlated with a conserved mutation at amino acid 200 in β-tubulin isotype 1. Mol. Biochem. Parasitol. *63*, 299–303.

Laing, R., Kikuchi, T., Martinelli, A., Tsai, I.J., Beech, R.N., Redman, E., Holroyd, N., Bartley, D.J., Beasley, H., Britton, C., et al. (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., Lecová, L., Skuce, P.J., Tait, A., and 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.

Le Jambre, L.F., Gill, J.H., Lenane, I.J., and Lacey, E. (1995). Characterisation of an avermectin resistant strain of Australian Haemonchus contortus. Int. J. Parasitol. *25*, 691–698.

Le Jambre, L.F., Gill, J.H., Lenane, I.J., and Baker, P. (2000). Inheritance of

avermectin resistance in Haemonchus contortus. Int. J. Parasitol. *30*, 105–111.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

de Lourdes Mottier, M., and Prichard, R.K. (2008). Genetic analysis of a relationship between macrocyclic lactone and benzimidazole anthelmintic selection on Haemonchus contortus. Pharmacogenet. Genomics *18*, 129–140.

Luo, X., Shi, X., Yuan, C., Ai, M., Ge, C., Hu, M., Feng, X., and 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.

Maciuca, S., del Ojo Elias, C., McVean, G., and Iqbal, Z. (2016). A natural encoding of genetic variation in a Burrows-Wheeler Transform to enable mapping and genome inference.

Nisbet, A.J., Meeusen, E.N., González, J.F., and Piedrafita, D.M. (2016). Immunity to Haemonchus contortus and Vaccine Development. In Advances in Parasitology, pp. 353–396.

Omura, S., and Crump, A. (2004). The life and times of ivermectin - a success story. Nat. Rev. Microbiol. *2*, 984–989.

Osei-Atweneboana, M.Y., Awadzi, K., Attah, S.K., Boakye, D.A., Gyapong, J.O., and Prichard, R.K. (2011). Phenotypic evidence of emerging ivermectin resistance in Onchocerca volvulus. PLoS Negl. Trop. Dis. *5*, e998.

Parts, L., Cubillos, F.A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S.J., Molin, M., Zia, A., Simpson, J.T., Quail, M.A., et al. (2011). Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res. *21*, 1131–1138.

Paten, B., Novak, A.M., Eizenga, J.M., and Garrison, E. (2017). Genome graphs and the evolution of genome inference. Genome Res. *27*, 665–676.

Rao, V.T.S., Siddiqui, S.Z., Prichard, R.K., and Forrester, S.G. (2009). A dopamine-gated ion channel (HcGGR3*) from Haemonchus contortus is expressed in the cervical papillae and is associated with macrocyclic lactone resistance. Mol. Biochem. Parasitol. *166*, 54–61.

Raza, A., Kopp, S.R., Bagnall, N.H., Jabbar, A., and Kotze, A.C. (2016). 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.

Redman, E., Sargison, N., Whitelaw, F., Jackson, F., Morrison, A., Bartley, D.J., and Gilleard, J.S. (2012). Introgression of ivermectin resistance genes into a susceptible Haemonchus contortus strain by multiple backcrossing. PLoS Pathog. *8*, e1002534.

Rezansoff, A.M., Laing, R., and 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.

Roos, M.H., Otsen, M., Hoekstra, R., Veenstra, J.G., and Lenstra, J.A. (2004). Genetic analysis of inbreeding of two strains of the parasitic nematode Haemonchus contortus. Int. J. Parasitol. *34*, 109–115.

Rose, H., Rinaldi, L., Bosco, A., Mavrot, F., de Waal, T., Skuce, P., Charlier, J., Torgerson, P.R., Hertzberg, H., Hendrickx, G., et al. (2015). Widespread anthelmintic resistance in European farmed ruminants: a systematic review. Vet. Rec. *176*, 546.

Rufener, L., Kaminsky, R., and Mäser, P. (2009). In vitro selection of Haemonchus contortus for benzimidazole resistance reveals a mutation at amino acid 198 of β-tubulin. Mol. Biochem. Parasitol. *168*, 120–122.

Sargison, N.D., Redman, E., Morrison, A.A., Bartley, D.J., Jackson, F., Gijzel, H.N., Holroyd, N., Berriman, M., Cotton, J.A., and Gilleard, J.S. (2018). A method for single pair mating in an obligate parasitic nematode. Int. J. Parasitol. *48*, 159–165.

Schwarz, E.M., Korhonen, P.K., Campbell, B.E., Young, N.D., Jex, A.R., Jabbar, A., Hall, R.S., Mondal, A., Howe, A.C., Pell, J., et al. (2013). The genome and developmental transcriptome of the strongylid nematode Haemonchus contortus. Genome Biol. *14*, R89.

Silvestre, A., and Cabaret, J. (2002). Mutation in position 167 of isotype 1 β-tubulin gene of Trichostrongylid nematodes: role in benzimidazole resistance? Mol. Biochem. Parasitol. *120*, 297–300.

Urdaneta-Marquez, L., Bae, S.H., Janukavicius, P., Beech, R., Dent, J., and 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.

Williamson, S.M., Storey, B., Howell, S., Harper, K.M., Kaplan, R.M., and 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.

van Wyk, J.A., and Malan, F.S. (1988). Resistance of field strains of Haemonchus contortus to ivermectin, closantel, rafoxanide and the benzimidazoles in South Africa. Vet. Rec. *123*, 226–228.

Xie, C., and Tammi, M.T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics *10*, 80.

Yadav, C.L., Kumar, R., Uppal, R.P., and Verma, S.P. (1995). Multiple anthelmintic resistance in Haemonchus contortus on a sheep farm in India. Vet. Parasitol. *60*, 355–360.