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 1EH, 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

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

32 **5. Department of Comparative Biology and Experimental Medicine, Faculty of Veterinary**
33 **Medicine, University of Calgary, Calgary, Alberta, Canada**
34 John S. Gilleard: jsgillea@ucalgary.ca
35 Elizabeth Redman: elmredma@ucalgary.ca
36 Andrew Rezansoff: amrezans@ucalgary.ca

37 **6. University of Edinburgh, Royal (Dick) School of Veterinary Studies, Edinburgh, EH25 9RG,**
38 **United Kingdom**
39 Neil Sargison: Neil.Sargison@ed.ac.uk

Abstract

Helminth infections cause an enormous disease burden in humans, livestock and companion animals. Helminth control is heavily dependent on the regular administration of anthelmintic drugs. Resistance to these drugs is widespread in many livestock systems, and there is concern that mass drug administration programs to control human helminths may be selecting for resistance. 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 that is a model for anthelmintic research. By combining genetic mapping and whole genome sequencing with novel analyses, 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 leading candidate genes. This work represents the first example of a major anthelmintic resistance locus being identified by genetic mapping in a parasitic nematode and represents the most comprehensive genetic analysis of this trait in a parasitic nematode to date.

57 **Introduction**

58 Helminths are an extremely diverse group of parasitic organisms, many of which are
59 responsible for significant morbidity and mortality in their hosts. Control of those
60 considered to be of human and veterinary importance is heavily dependent on the
61 large-scale administration of anthelmintic drugs. Anthelmintic drugs such as the
62 macrocyclic lactone, ivermectin, have been extremely successful in the control of a
63 number of helminths in both humans and animals (Omura and Crump, 2004). The
64 importance of the discovery and development of this drug as an anthelmintic was
65 recently recognised by the award of the 2015 Nobel Prize for Medicine or Physiology.
66 However, these successes are now threatened by the emergence of drug resistance.
67 Anthelmintic drug resistance is already widespread in many species of parasitic
68 nematodes of livestock and rapid acquisition of resistance has been documented
69 against single and multiple drug classes (Kaplan and Vidyashankar, 2012; Rose et
70 al., 2015), and is already a major problem for global agricultural production and
71 animal welfare. There are growing concerns regarding reduced efficacy of
72 compounds used in mass drug administration (MDA) programs for some human
73 helminths, which could herald the emergence of resistance (Crellen et al., 2016;
74 Doyle et al., 2017; Geerts and Gryseels, 2000; Osei-Atweneboana et al., 2011).
75 Consequently, anthelmintic drug resistance is an emerging threat to the sustainable
76 control of human parasites, particularly those targeted by MDA programs. However,
77 in spite of the importance of this issue, remarkably little is known regarding the
78 molecular mechanisms of resistance to most anthelmintic drug groups, with the
79 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 parasite of wild and domesticated ruminants of global distribution, and causes major health and economic impacts in sheep and goats worldwide. This parasite has a propensity to develop resistance to anthelmintic drugs, with resistance to almost all of the classes of anthelmintic drugs having 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). *H. contortus* isolates 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* is emerging 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 genetic information about the genome structure and transcriptional differentiation of the major life stages (Doyle et al., 2018; Laing et al., 2013; Schwarz et al., 2013).

101 The genetic characterisation of drug resistance in parasitic nematodes offers the
102 promise of sensitive and specific diagnostic tools for use in research, surveillance
103 and clinical diagnostics to support sustainable parasite management programs.
104 Understanding the mechanisms of resistance to current drugs should also provide
105 important information to help direct the development of new anthelmintic drugs
106 (Hefnawy et al., 2017). However, the identification of genetic loci associated with
107 drug resistance in these organisms remains a significant challenge. The only
108 anthelmintic drug class to date for which the major resistance loci have been
109 unequivocally identified is the benzimidazoles, where three mutations in the the
110 isotype-1 ß-tubulin gene of *H. contortus* and several other trichostrongylid
111 nematodes species are the major causal loci (Elard and Humbert, 1999; Kwa et al.,
112 1994; Rufener et al., 2009; Silvestre and Cabaret, 2002). In the case of other
113 anthelmintic drug classes, the genetic mediators of resistance remain obscure; for
114 example, mutagenesis experiments in the free-living nematode *Caenorhabditis*
115 *elegans* have identified many genetic loci capable of conferring low levels of
116 ivermectin resistance, however, very few mutations confer high levels of resistance
117 (Hunt, 1995). Concurrent mutation of three genes, *avr-14*, *avr-15* and *glc-1,* each
118 encoding a glutamate-gated chloride channel (GluCl), confers a high level of
119 resistance to ivermectin (Dent et al., 2000); however, mutations in orthologous GluCl
120 genes have not been associated with ivermectin resistance in parasitic nematodes.
121 In the case of parasitic nematodes, largely due to the lack of genomic resources,
122 research into the genetic basis of ivermectin resistance has been dominated by the
123 examination of small numbers of candidates genes (Kotze et al., 2014)). Candidate
124 genes are generally chosen based on their potential roles as drug targets or in drug

efflux and examined for associations with sequence polymorphisms with the resistance phenotype. 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.

Genome-wide and genetic mapping approaches are at last becoming feasible for *H. contortus* due to progress in genetic crossing methodologies (Sargison et al., 2018) and an increasing complement of genomic resources. Previous work has undertaken two serial backcrosses 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**). These successfully introgressed ivermectin resistance loci from the two resistant strains into the susceptible MHco3(ISE) genomic background. Genetic validation of these backcross experiments, using a panel of microsatellite markers, revealed one such marker, Hcms8a20, to be present in an introgressed region of the genome. This was confirmed by subsequent work that further showed that none of the other 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*,

147 offers an opportunity to characterise precisely the genome-wide evidence of
148 introgression and selection surrounding the Hcms8a20 locus.

149 Here, we build on these two previous genetic crosses, extending them for further
150 generations of passage with ivermectin selection (**Figure 1 C**), and generating whole
151 genome sequencing data from four steps in the experiment. We analyse these data
152 with novel statistical methods and exploit a chromosome-scale genome assembly to
153 map ivermectin resistance in the two genetically and geographically distinct strains of
154 *H. contortus*. Previous research has highlighted the value of novel methods in
155 population genetics for analysing experimental cross populations (Illingworth and
156 Mustonen, 2013). Here, in order to gain the maximum possible insight into the data
157 collected we developed a bespoke analysis, extending previous statistical work
158 (Abkallo et al., 2017; Illingworth et al., 2012), to analyse data from the experiment in
159 a manner that explicitly accounts for the genetic structure of the organisms in the
160 experiment; this approach was combined with more traditional population genetic
161 tools. Together, our methods provided a consistent and deep insight into the data,
162 identifying a single region of genomic introgression under selection for resistance.
163 This work represents the most comprehensive analysis of the genetics of
164 anthelmintic resistance in a parasitic nematode to date and demonstrates the power
165 of genetic crossing and contiguous genome assembly to eliminate false positive
166 genetic signals typically linked to resistance.

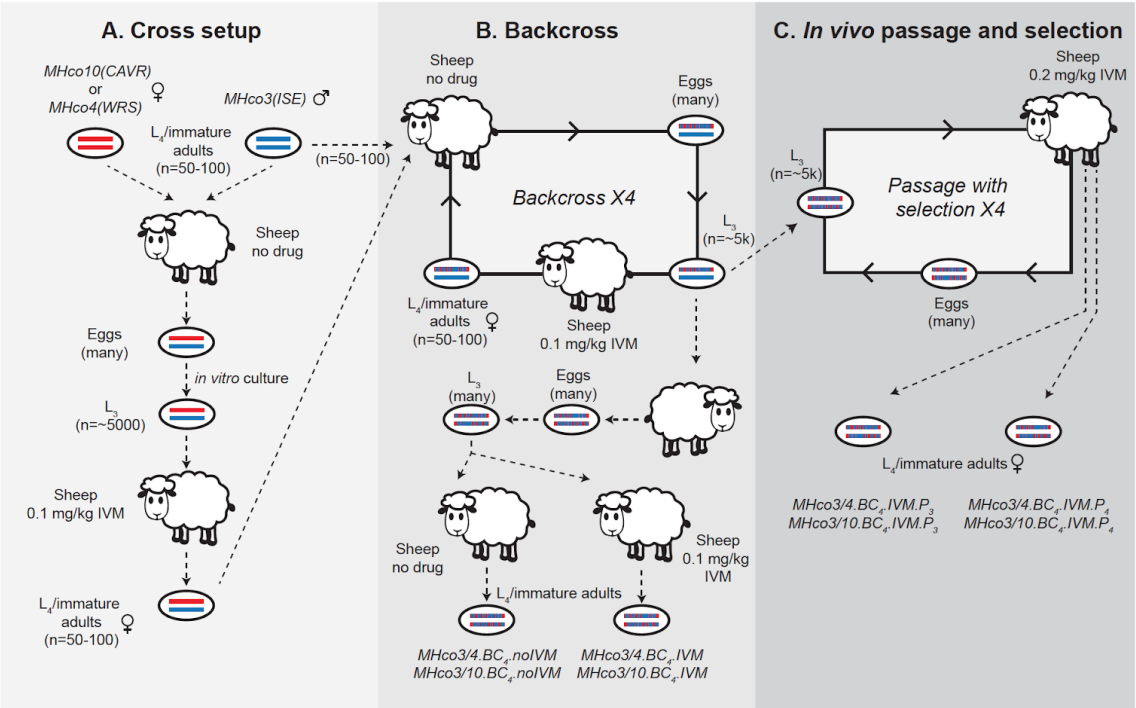


Figure 1. Outline of the initial crosses, backcrosses, *in vivo* passage and selection experiments. A. Ivermectin resistant strains MHco10(CAVR) or MHco4(WRS) ("resistant" haplotypes are depicted as red lines) were crossed with an ivermectin sensitive strain MHco3(ISE) ("susceptible" haplotypes as blue 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

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

192 **Figure Supplement**
193 **Table S1.** Sequencing datasets used in this study

194 **Results**

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

196 We have taken a genome-wide approach to characterise genetic diversity within and
197 between two ivermectin resistant strains (MHco10[CAVR] and MHco4[WRS]) and
198 one ivermectin susceptible (MHco3[ISE]) strain of *H. contortus*. This study exploits a
199 greatly improved, chromosomal-scale *H. contortus* reference genome sequence
200 (**Figure 2A**), which was derived from sequencing and assembly of an inbred isolate
201 of the MHco3(ISE) strain used here. Whole genome sequencing of pooled worms
202 from each of the geographically distinct strains (**Figure 2B**) revealed abundant levels
203 of nucleotide diversity throughout the five autosomes of the genome, with almost
204 twice as much diversity on average in the resistant strains (MHco10[CAVR] mean π
205 = 0.035 ± 0.008 standard deviations (SD); MHco4[WRS] mean π = 0.038 ± 0.008
206 SD) relative to the susceptible strain (MHco3[ISE] mean π = 0.022 ± 0.008 SD)
207 (**Figure 2C; Figure S1A**). In contrast to the autosomes, the X chromosomal
208 scaffolds had significantly less diversity in all three strains analysed (MHco3[ISE]
209 mean π = 0.008 ± 0.008, MHco10[CAVR] mean π = 0.014 ± 0.012, MHco4[WRS]
210 mean π = 0.017 ± 0.014) (**Figure 2C; Figure S1A**). This is consistent with a recent
211 genome-wide analysis that revealed the X chromosome contained as little as 10% of
212 the genetic diversity relative to the autosomes (Doyle et al., 2018). The difference
213 between sensitive and resistant strains is perhaps not surprising given their history;
214 MHco3(ISE) is a partially inbred laboratory strain that has been maintained in a
215 closed, experimental flock, whereas the resistant strains have not been subjected to
216 experimental inbreeding and have been isolated from outbred populations more
217 recently.

218 We examined short-range haplotype diversity in each strain by measuring linkage
219 disequilibrium (LD) between pairs of SNPs detected in paired reads (**Figure S1B**).
220 Although restricted by the poolseq design to pairwise SNP comparisons less than
221 500 bp apart, i.e., within a read pair, we nevertheless observed considerable decay
222 in LD over this distance. Moreover, the rate of LD decay was correlated with
223 nucleotide diversity, such that there was a greatest loss of LD in MHco4(WRS)
224 followed by MHco10(CAVR) over the 500 bp, and a greater proportion of variants in
225 LD in the less diverse, susceptible MHco3(ISE) strain.

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

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

239 significant variation between strains (**Figure S3 A,C,E**), with a large number of
240 deletions, and substantially fewer duplications and inversions detected (**Figure S3**
241 **B,D,F**).

242 The genetic diversity between each of the three strains was assessed by pairwise
243 analysis of genetic diversity (measured by F_{ST} from single nucleotide variants in 10
244 kb windows), which confirmed significant genome-wide genetic diversity and strain
245 differentiation (**Figure 2D**). These data are consistent with the hypothesis that this
246 diversity reflects independent evolutionary histories due to being geographically and
247 therefore reproductively isolated (**Figure 2B**), and in turn most of this genetic
248 variation reflects genetic structure between these populations. These data
249 emphasise the challenge of characterising genetic variation associated with
250 phenotypes such as drug resistance by simply comparing genetic diversity between
251 a susceptible and resistant strain without accounting for this background genetic
252 variation.

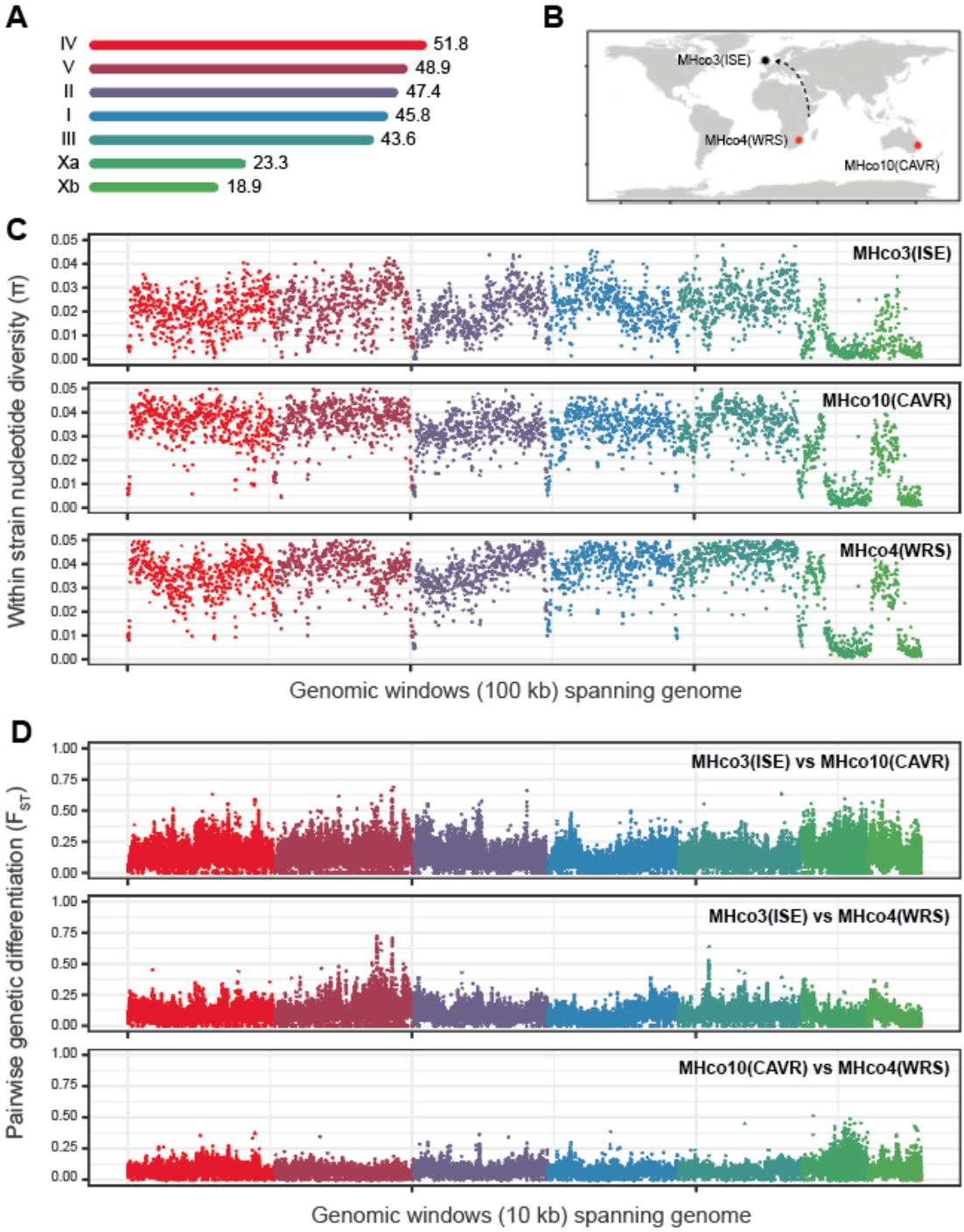


Figure 2. Within and between strain genetic diversity of parental isolates used in the genetic cross. A. The 279 Mb genome assembly of *H. contortus* consists of five autosomal and two X chromosome scaffolds, named based on synteny with *Caenorhabditis elegans* chromosomes. The size of each scaffold 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

To utilise the genetic variation that differentiates these strains as described above, a backcross experiment was 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). Here, we describe whole genome pairwise comparisons of genetic diversity throughout the backcross to assess the degree of introgression of resistance allele(s) into the susceptible genetic background, and to detect evidence of selection associated with resistance (**Figure 3**). Each pairwise comparison made 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 (MHco3/10.BC_4.noIVM & MHco3/4.BC_4.noIVM), there would be relatively little genetic differentiation across the whole genome between the population and MHco3(ISE). By contrast, where populations were subjected to further ivermectin treatment, we hypothesised that regions of the genome would have variable levels of

genetic differentiation with MHco3(ISE), with high levels of differentiation in regions of the genome linked to an ivermectin resistance conferring locus and low levels of differentiation elsewhere.

In both crosses, as we hypothesised, there was 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). This is despite the fact that resistant alleles had been selected for in each round of the backcross; post backcross, susceptible alleles would be segregating at a frequency of ~0.97 due to the repeated backcrossing with MHco3(ISE), and therefore largely indistinguishable from the susceptible parent. Upon selection (**Figure 3**; MHco3/10.BC_4.IVM and MHco3/4.BC_4.IVM), we observed a region of differentiation from the susceptible parent, particularly in the MHco3/10 cross, 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 was observed to show both a marked and progressive increase in differentiation throughout the crosses. To test this more explicitly, we regressed F_{ST} per window throughout the genome over the course of the cross; only the chromosome V region demonstrated consistent change associated with introgression over time (**Figure S4 A,B**).

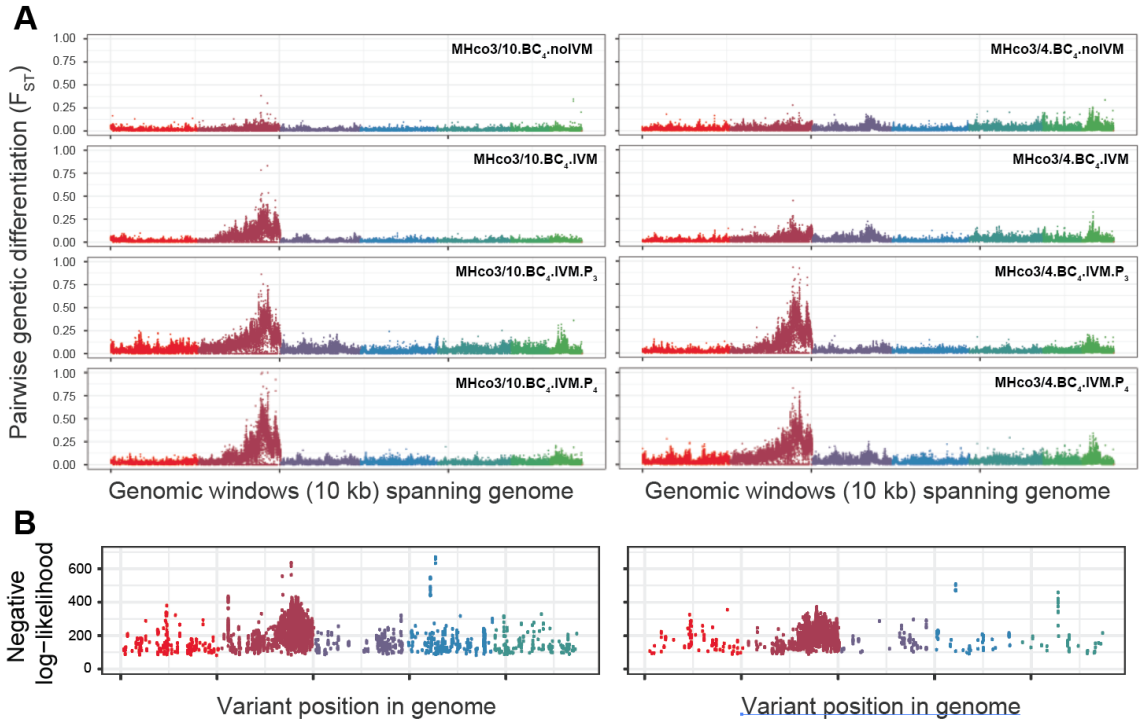


Figure 3. Genome-wide analysis of ivermectin associated loci. A. Pairwise genetic diversity throughout the MHco3/10 (left plots) and MHco3/4 (right plots) backcrosses. In both crosses and at each point in the backcross, the experimental strain was backcrossed against the MHco3(ISE) parental strain – the data is therefore presented to compare the genetic diversity (F_{ST} measure in 10 kb windows using *popoolation2*) between the MHco3(ISE) and 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.

These data also confirmed that the manner in which the population changed at a given site was highly correlated between the two crosses (**Figure S4C**). Collectively,

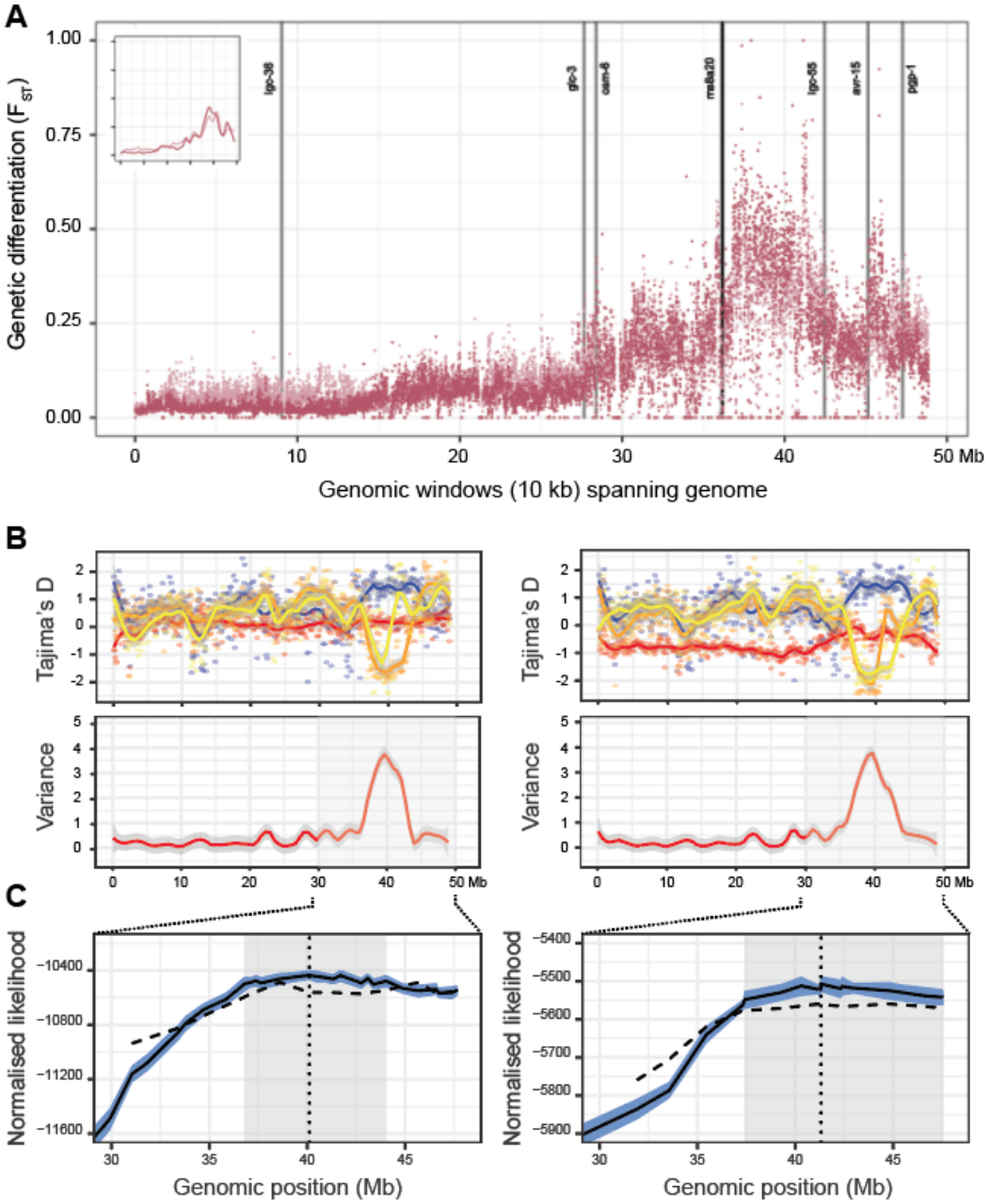
these data suggest that there is a single introgression region on chromosome V that is under selection upon drug exposure, and that this region is under selection in the two independent serial backcrosses performed. This conclusion was supported by an independent evolutionary analysis, which compared 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**). Alleles with frequencies that were strongly inconsistent with the neutral expectation in all four of the collected samples, suggesting that they had potentially evolved under the influence of selection, were over-represented in chromosome V (**Figure 3B**), with 71% and 91% of the total number of such alleles being found in this region for the MHco3/10 (**Figure S6 A)** and MHco3/4 (**Figure S6 B)** crosses, respectively. This supports the inference of a single selected region in the parasite genome towards the right hand end of chromosome V.

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, with a second but lesser increase in genetic differentiation between 45-48 Mb that is more prominent in the MHco3/10 cross but less so in the MHco3/4 cross (**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

346 shown to be genetically linked to ivermectin resistance in the preliminary analysis of
347 these backcrosses (Redman et al., 2012).

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



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

369 (MHco3/10.BC_4.IVM.P_4 or MHco3/4.BC_4.IVM.P_4; yellow) of the crosses. Tajima's D
370 was calculated using *npstat* in 100 kb windows spanning the genome. The variance
371 in the mean value of Tajima's D among MHco3(ISE) and passages 3 and 4 samples
372 is presented as smoothed line (red). The null expectation is that variance between
373 these samples will be low in regions of the genome not under selection. **C**. Inferred
374 likelihoods from the multi-locus population genetic model. The mean likelihood of the
375 best fitting single-locus model at each locus position is shown by the solid black line;
376 the light blue interval around this line shows the 5-95% confidence interval for this
377 statistic calculated by a bootstrapping process. The position of the maximum
378 likelihood value of the likelihood is shown by the vertical black dotted line; a
379 confidence interval for this position, calculated from the bootstrapping values, is
380 shaded in gray. The mean likelihood of the best fitting two-locus model at including
381 each locus position is shown by the black dashed line.

382 **Figure supplements**

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

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

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

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

389 A multi-locus population genetic model assuming the presence of a single selected
390 allele in the introgression region inferred the selected resistance conferring allele to
391 be in a location consistent with the previous analysis. In the first step of this model,
392 putative segregating sites in chromosome V between the resistant and susceptible
393 parental populations were identified; a total of 474 such sites were identified in the
394 MHco3/10 dataset, with 157 sites in the MHco3/4 dataset; calculations were
395 performed upon data from these loci. A maximum likelihood model inferred the
396 location of the driver to be in a broadly consistent location in each experiment, with
397 peaks at 40.10 and 41.31 Mb respectively. However, there was considerable
398 uncertainty in the location of the selected allele, with confidence intervals of 36.86 to

399 44.03 Mb and 37.41 to 47.56 Mb in each case (**Figure 4 C**). These broad confidence
400 intervals were explained by the population structures inferred by the model, which
401 explicitly considered the location of crossover recombination events between the
402 parental strains throughout the experiment. While the precise location of crossover
403 events is stochastic, repeated calculations produced an ensemble of replicate
404 population structures for each set of model parameters. Example structures
405 generated using the maximum likelihood parameters for each population show how
406 selection for resistance drives the accumulation of genetic material from the resistant
407 parent through the course of the experiment (**Figure 5**); these outputs may be
408 contrasted with equivalent data generated under a model of selective neutrality
409 (**Figure S10**). Across 250 replicates, a mean of 177.1 crossover recombination
410 events were predicted in the final CAVR population (range 120 to 224), giving a
411 mean length of a block of parental genome of 17.64 Mb; in the final WRS population
412 a mean of 186.4 recombination events were seen (range 143 to 238), giving a mean
413 length of a block of parental genome of 17.06 Mb. Under selection, blocks of genome
414 containing the selected allele are favoured; the large size of each block reduces the
415 precision with which the location of the allele under selection can be identified.

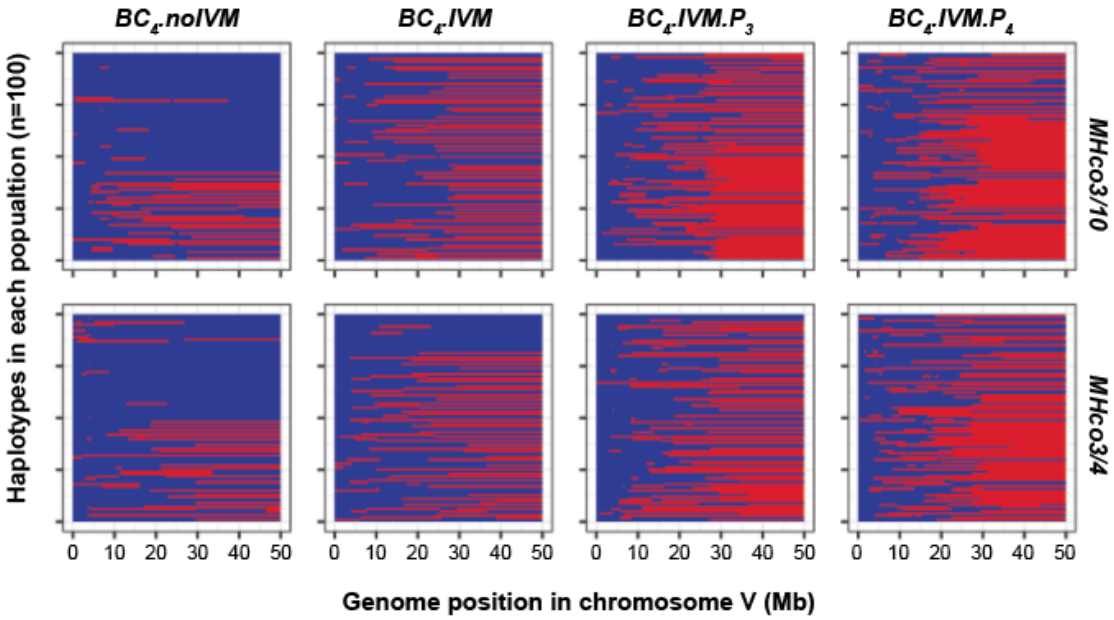


Figure 5. Haplotype structure of chromosome V in an example output from the multi-locus model. Segments of genome from the parent containing the resistance allele are shown in red, while segments of genome from the susceptible parent are shown in blue. Data shown were generated using the maximum likelihood 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 large region of introgression was expected due to limited extent of recombination in the population. To test this, we investigated an extended multi-locus model incorporating selection on alleles at two loci within the introgression region, as the presence of multiple drivers could explain the second, smaller distinct peak of differentiation 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 these alleles must be at least 2 Mb apart in the genome. Under this constraint, the best fitting model identified

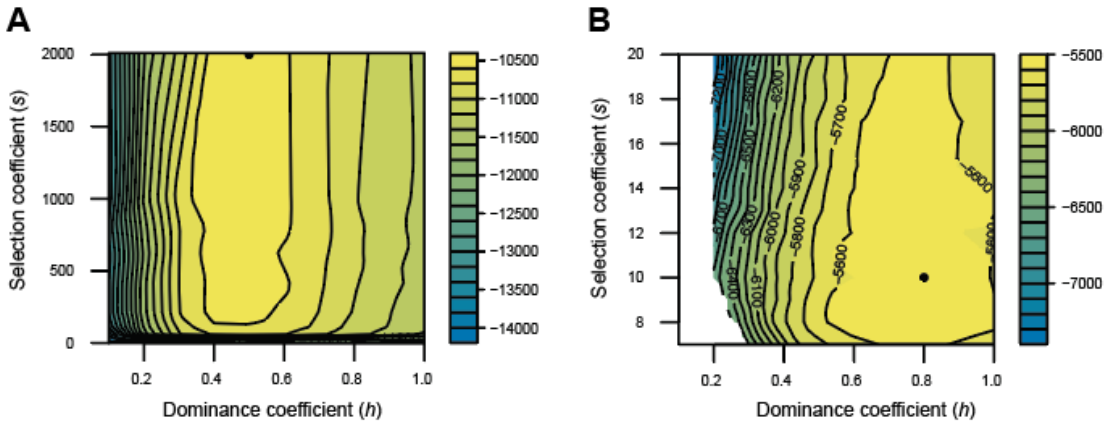
432 variants under selection at the positions 38.59 Mb and 45.68 Mb in the MHco3/10
433 data (**Figure S8 A)**, and at the positions 41.20 Mb and 44.80 Mb in the MHco3/4
434 data (**Figure S8 B**). However, under the imposed constraint, the two-driver model
435 performed worse than the single-driver model shown above, favouring an
436 explanation in which only one variant is under selection. (We note that removing the
437 constraint that selected alleles be separated would guarantee a likelihood at least
438 equal to that of the single-driver model). Due to the limited extent of recombination in
439 the experiment, we cannot rule out the presence of more than one drug-resistance
440 variant in the identified region of chromosome V, but we did not find evidence in
441 favour of a second selected allele.

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

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

460 **Strength of selection**

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



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

484 **Discussion**

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

487 Given the importance of ivermectin, and other macrocyclic lactones, to human and
488 animal health, the widespread emergence of resistance in multiple parasite species
489 is a major concern. In spite of several decades of research, the molecular genetic
490 basis of ivermectin resistance in parasitic nematodes has remained elusive. This is
491 due to a combination of factors including a lack of genomic resources, the general
492 experimental intractability of parasitic nematodes and the often high levels of genetic
493 variation which complicates genetic analysis. *H. contortus* is the most experimentally
494 tractable and most widely accepted parasite model for anthelmintic resistance
495 research (Gilleard, 2013). Many candidate genes have been proposed to be
496 associated with ivermectin resistance in *H. contortus*, chosen largely on *a priori*
497 assumptions regarding their known or predicted function. Such candidate genes
498 have typically been implicated as involved in ivermectin resistance largely on the
499 basis of studies comparing SNP or haplotype frequencies between small numbers of
500 resistance and susceptible isolates (Blackhall et al. 1998; Eng et al. 2006; Luo et al.
501 2017; Blackhall et al. 2003; Urdaneta-Marquez et al. 2014) or pre- and post-
502 ivermectin treatment (Raza et al. 2016; Lloberas et al. 2013). However, careful
503 validation of a number of these candidates in different field populations (Laing et al.,
504 2016) or in controlled crosses (Rezansoff et al., 2016) have failed to support the
505 initial associations observed. The discordance among studies reflects the difficulty in
506 identifying real associations between genotype and phenotype simply by comparing
507 resistant and susceptible populations, or individuals, due to the confounding effects

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

515 In this paper, we describe the analysis of whole genome sequence data collected
516 from the most recent phase of a multi-generational crossing experiment conducted in
517 populations of the parasitic worm *H. contortus*. Our analysis of genome sequence
518 data unequivocally identifies a single region within chromosome V as containing at
519 least one major ivermectin resistance allele. This major ivermectin resistance locus is
520 common to two independent ivermectin strains - MHco10(CAVR) and MHco4(WRS) -
521 that were originally isolated from different continents; Australia and Africa
522 respectively. This genomic region is clearly distinguishable in the experimental data,
523 being highlighted by each of the different population genetic methods applied to our
524 data. There are no other comparable regions of introgression elsewhere in the
525 genome in either of the two backcross experiments suggesting this is the most
526 important ivermectin resistance locus in the MHco10(CAVR) and MHco4(WRS)
527 strains. There has been much discussion regarding the likely multigenic basis of
528 ivermectin resistance but, although there could be additional additive loci, these
529 results strongly suggest a single locus (or potentially multiple closely linked loci) is
530 likely to be the major effector of ivermectin resistance in these two strains.

531 While each of our analytical methods gave a consistent identification of the region
532 under selection, our multi-locus population genetic analysis provided an estimate of
533 the uncertainty in the location of the selected allele, inferring its position to lie at a
534 position between 37 Mb and 45 Mb. While the region in which we infer a selected
535 allele to exist is large, our result nevertheless provides some ability to discriminate
536 between potential ivermectin resistance alleles in the worm population. Most of the
537 leading published candidate ivermectin resistance genes including *Hco-glc-5*,
538 *Hco-avr-14*, *Hco-pgp-9*, *Hco-lgc-37*, *Hco-pgp-2* and *Hco-dyf-7* were not located in or
539 near the region of introgression, which is consistent with a recent study using
540 targeted sequencing of these genes in which none of them showed a signal of
541 introgression in these two backcross experiments (Rezansoff 2016). This suggests
542 that none of these candidate genes are a major ivermectin resistance-conferring
543 locus in the MHco10(CAVR) and MHco4(WRS) strains. Although, we identified three
544 other previously described candidate genes - *Hco-lgc-55*, *Hco-avr-15*, and
545 *Hco-pgp-1* - adjacent to the introgression region, none were found within the peak of
546 the region suggesting that these also are unlikely to be major direct targets of
547 ivermectin-mediated selection. We acknowledge that we cannot discount previously
548 described candidate genes as mediators of resistance in other isolates of *H.*
549 *contortus*, or in other nematode species. Moreover, our data does not exclude a
550 driver of transcriptional regulation within the introgression region that is under
551 selection, which itself influences expression of other genes, including candidate
552 genes, outside of the introgression region.

553 What might account for the large number of candidate genes previously suggested
554 to be associated with ivermectin resistance? The answer might lie in our direct
555 comparison of genetic diversity between the parental strains (**Figure 2D**). This
556 visualisation of the extent of genetic divergence between these strains makes it
557 easier to understand how particular sequence polymorphisms might be naively
558 attributed to being associated with resistance when in fact they simply represent one
559 of many strain specific genetic variations that occur as a result of the independent
560 evolutionary histories and lack of interbreeding between the strains. This highlights
561 the challenges of interpreting simple direct comparison of strains, even with genomic
562 approaches, when trying to disentangle those genetic polymorphisms underlying
563 resistance from a complex background of genetic variation independent of a strains
564 resistance status. The use of controlled genetic crosses as presented here and
565 elsewhere (Choi et al., 2017; Redman et al., 2012; Rezansoff et al., 2016), whereby
566 population genetic structure can be explicitly accounted for in the analysis, provides
567 a powerful way to mitigate the challenge of discriminating resistance causing alleles
568 from background genetic variation. As such, the locus we have identified on
569 chromosome V represents the only unequivocal major ivermectin resistance locus
570 identified in any parasitic nematode species to date.

571 **Population genetic and evolutionary modelling helps to statistically define the**
572 **boundaries of the genetically introgressed genomic region linked to the**
573 **ivermectin resistance conferring locus**

574 Inference of selection from the multi-locus model suggested the presence of stronger
575 selection for the drug resistance allele in the MHco3/10 cross than in the MHco3/4

576 cross (**Figure 6**). Additionally, although the data suggests that resistance is at least
577 additive in both crosses, i.e., heterozygotes are likely to confer some resistance, but
578 increased resistance is achieved by having two resistant alleles, there is some
579 evidence for dominance in the MHco4(WRS) strain, whereby the second resistant
580 allele confers a lesser fitness advantage. A previous study reported resistance to be
581 a completely dominant trait in the CAVR strain (Le Jambre et al., 2000) on the basis
582 that no difference was observed in the resistance levels of heterozygote and
583 homozygote resistance worms. Our finding of additivity arises from the evolution of
584 the worm populations during passage following the backcross experiment (i.e., in
585 BC_4.IVM.$P_{3/4}$ samples versus the BC_4.IVM data), and particularly the fixation of
586 alleles from the resistant parent at putative segregating sites near the inferred
587 position of resistance. In our model, alleles can reach fixation only where there is
588 selection in favour of the homozygote resistant type over and above that for
589 heterozygote resistant individuals.

590 We have here utilised a range of population genetic methods to attempt to localise
591 selection. In addition to the standard metrics of F_{ST} and Tajima's D, we have
592 implemented single-locus and multi-locus methods which explicitly account for the
593 history of the cross population. Each approach has limitations. While most models
594 neglect linkage between adjacent sites in the genome, our multi-locus model
595 explicitly accounts for genomic structure during the cross. However, even this model
596 involves the use of heuristic measures to identify segregating sites between the two
597 parental strains, and is based upon a map of recombination in the worm genome
598 (Doyle et al., 2018) that, while the state of the art for parasitic species, was derived

599 from a limited number of observed crossover events. None of our models accounts
600 for genomic rearrangements during the experiment, which are inherently difficult to
601 observe via short read sequencing.

602 The output of our population genetic model suggests that our ability to resolve the
603 precise location of the variant under selection is limited by the number of
604 recombination events occurring in the worm population. Precise identification of the
605 location of an allele under selection requires that the allele not be linked to other
606 nearby alleles in at least some fraction of individuals in the population. While the
607 inherent biological recombination rate here has a role, some of the lack of
608 recombination events in this case results from bottlenecking of the population during
609 the course of the experiment due to the limited number of L_4 worms that it was
610 possible to collect from each sheep and transplant in the successive generation of
611 the cross. The limited individuals, and hence, limited recombinants, reduced the
612 resolution of the experiment to resolve selection. Genetic drift induced by successive
613 population bottlenecks introduces considerable uncertainty in the outcome of the
614 experiment, such that replicate sets of allele frequencies from our model showed
615 considerable differences between each other (**Figure S5)**. This implies that the data
616 from the experiment itself should be understood as the output of a stochastic
617 process; beyond the clear large-scale patterns observable in the data and detailed
618 above, more minor details of the output might not be seen again were the experiment
619 to be repeated.

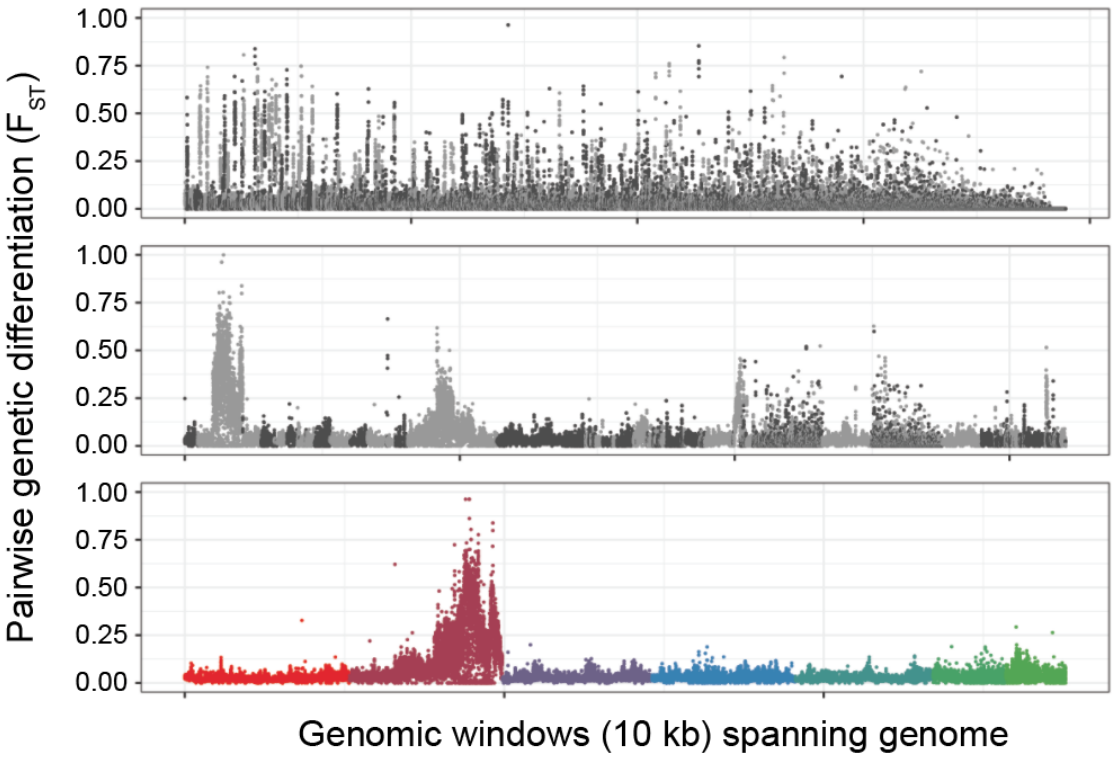
620 The structure of the experiment thus imposes a limit on our ability to infer the location
621 of a selected allele, and an improved characterisation of selection would likely best
622 be achieved through the use of alternative experimental procedures. Conducting
623 further generations of cross as performed within this experimental framework would
624 induce more recombination events, reducing the mean size of parental genomic
625 blocks. However, such a course of action would be limited in scope were it not
626 accompanied by the use of larger L_4 populations, for example by the simultaneous
627 passage of the population through multiple animals in parallel. Clearly, a larger or
628 longer experiment would have cost and welfare implications: the statistical
629 framework we have developed will help design and justify such an approach.

630 **The importance of chromosomal-scale genome assemblies and the dangers of**
631 **using fragmented draft genomes in genome-wide genetic mapping and**
632 **population genomic analyses**

633 The success of identifying a single region of introgression was dependent upon a
634 greatly improved and highly contiguous, chromosomal-scale reference genome
635 assembly for the *H. contortus* isolate, MHco3(ISE).N1. This isolate was used in the
636 original draft assembly of this species (Laing et al., 2013), which was derived via
637 inbreeding of the same MHco3(ISE) strain used in the backcrosses presented. In the
638 absence of a contiguous chromosomal scale genome assembly, interpretation of
639 these analyses can be extremely challenging; true genetic signal(s) can be obscured
640 by technical artifacts associated with a fragmented assembly that can include short
641 contigs lacking spatial orientation, multiple haplotypes present, collapsed paralogs,

642 and poor resolution of repeat structures. As a consequence, read mapping and
643 variant calling in fragmented genomes can be suboptimal. These technical
644 challenges are exacerbated in highly polymorphic species such as *H. contortus* and
645 *Teladorsagia circumcincta* (Choi et al., 2017), in which reference genomes were
646 constructed from pools of individuals that, despite inbreeding, still contained
647 significant genetic variation. Perhaps the biggest problem with using a fragmented
648 genome assembly for genome-wide analyses is the inability to determine linkage.
649 Signals of selection will often be dispersed across multiple scaffolds of a fragmented
650 draft genome assembly when in reality they are co-located in a single genomic
651 region. This can potentially can give rise to the erroneous conclusion that multiple
652 independent loci show a signal of selection when it may be that only a single
653 selected locus exists. For example, a major effect single locus that has a physically
654 extensive signature of selection due to a hard selective sweep, and the associated
655 hitchhiking effect, could easily be interpreted as multiple independent unlinked loci
656 under selection. These artefacts are evident when examining the sequence data we
657 have generated on different versions of the *H. contortus* MHco3(ISE).N1 reference
658 genome assembly. In this case, the introgression region of chromosome V only
659 became apparent after significant improvement in chromosome contiguity and once
660 chromosome V was fully assembled (**Figure 7**). In the earlier, more fragmented, draft
661 genome assemblies, the signature of selection was dispersed across multiple
662 scaffolds making the number of major loci underlying ivermectin resistance in the
663 MHco10(CAVR) and MHco4(WRS) strains unclear. This will be an increasingly
664 important issue as unfinished draft reference genomes are produced for an
665 increasing number of non-model species and whole genome sequencing becomes

666 cheaper and more widely available. It will be important for those using genome-wide
667 data for population genomic and genetic studies to be aware of the limitations of
668 using unfinished draft reference genomes in various genome-wide analyses.



669 **Figure 7: Demonstrable effect of genome contiguity on the resolution of the**
670 **introgression region.** The same pairwise comparison - MHco3(ISE) *vs*
671 MHco3/10.BC_4.P_4 - is presented using on three versions of the *H. contortus* genome;
672 the published V1 draft assembly (Laing et al., 2013) (top; N50 = 0.083 Mb, N50(n) =
673 1,151, n = 23,860), an intermediate improved assembly (middle; N50 = 5.2 Mb,
674 N50(n) = 16, n = 6668), and the chromosomal-scale assembly presented here and
675 elsewhere (Doyle et al., 2018)(bottom; N50 = 47.4 Mb, N50(n) = 3, n = 8). The top
676 and middle plots are coloured light and dark grey to reflect alternate contigs /
677 scaffolds, whereas the bottom plot is coloured as in Figure 2A.

678 Our comprehensive genome-wide analysis of parental strain genetic diversity
679 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 such as Pacbio or Oxford Nanopore sequencing, which will allow for 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 previous 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.

717 **Acknowledgement**

718 Work at the Wellcome Sanger Institute was funded by Wellcome (grants 098051 and
719 206194) and by the Biotechnology and Biological Sciences Research Council
720 (BB/M003949). CJRI was supported by a Sir Henry Dale Fellowship, jointly funded
721 by Wellcome and the Royal Society (grant 101239/Z/13/Z). We thank Pathogen
722 Informatics and DNA Pipelines (WSI) for their support and expertise, and Guillaume
723 Sallé for constructive feedback on the manuscript.

724 **Competing interests**

725 The authors declare no competing interests.

726 **Materials and methods**

727 **Background of strains and original backcross**

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

737 The experiment described here extends two previously-described backcross
738 experiments. The construction, phenotypic validation, and initial microsatellite
739 analysis of these backcrosses has been described previously (Redman et al., 2012),
740 and is outlined in **Figure 1A and B.** Briefly, two independant crosses were performed
741 in parallel: (i) MHco3(ISE) x MHco10(CAVR), and (ii) MHco3(ISE) x MHco4(WRS). In
742 the first generation of each cross, female worms of the ivermectin resistant parental
743 strains [MHco10(CAVR) or MHco4(WRS)] were crossed with male MHco3(ISE)
744 worms, to generate lines designated MHco3/4 and MHco3/10, respectively (**Figure**
745 **1A**). After the first cross, the F_1 generation females were mated to male MHco3(ISE),
746 resulting in backcross generations designated MHco3/10.BC_n and MHco3/4.BC_n,
747 where n is the number of backcross generations (BC). This involved the recovery of
748 immature worms from the abomasum at day 14 post infection by necropsy of donor

749 sheep, which were washed in physiological saline (0.85% NaCl) after which worm
750 sex was determined (Denham, 1969), followed by surgical transfer of 45-100 male
751 MHco3(ISE) and 50-100 female F_1 L_4/immature adult worms into worm-free recipient
752 lambs all within 2 h of the original collection. This process of backcrossing was
753 repeated for a total of 4 generations, i.e., MHco3/10.BC_4 and MHco3/4.BC_4 (**Figure**
754 **1B**). *In vivo* ivermectin selection was applied after mating and before collection of
755 eggs to enrich for ivermectin resistant adults to be used in the subsequent round of
756 the backcross. A controlled efficacy test was performed on the three adult strains
757 and MHco3/4.BC_4 and MHco3/10.BC_4 strains to determine the initial levels of
758 ivermectin efficacy, and the degree of resistance acquired in the introgressed lines.
759 After four rounds of backcrossing, a selection experiment was performed.
760 MHco3/10.BC_4 and MHco3/4.BC_4 strains were used to infect recipient sheep, after
761 which eggs were collected and in vitro culture to L_3. These infective larvae were used
762 to infect two sheep per strain, one that received 0.1 mg/kg ivermectin and one that
763 remained drug naive. At 7 days post treatment, treated and naive L_4/immature adults
764 were recovered by necropy and stored for molecular analysis.

765 The backcrosses described above were extended by performing a further four
766 rounds of *in vivo* ivermectin selection, during which mating within the BC_4 population
767 continued (**Figure 1C**). For each cross, progeny of the BC_4 generation were cultured
768 to L_3 and used to infect a donor sheep, which was subsequently treated with
769 ivermectin (0.1 mg/kg in round 1, followed by 0.2 mg/kg in following rounds). Eggs
770 from ivermectin-treated survivor adults were collected post-treatment, cultured to L_3,
771 and used to infect a new donor sheep. L_4/immature adults were collected by

772 necropsy after rounds three (i.e., BC_4.IVM.P_3) and four (i.e., BC_4.IVM.P_4) of selection
773 and stored for molecular analysis.

774 **Library preparation and sequencing**

775 Whole genome sequencing was performed on the three parental strains
776 (MHco3(ISE), MHco10(CAVR) & MHco4(WRS)), and from each cross, pre and post
777 ivermectin treatment following four rounds of backcrossing (MHco3/10.BC_4.noIVM,
778 MHco3/4.BC_4.noIVM, MHco3/10.BC_4.IVM, MHco3/4.BC_4.IVM), and after 3 and 4
779 rounds of additional selection after the backcross (MHco3/10.BC_4.IVM.P_3,
780 MHco3/4.BC_4.IVM.P_3, MHco3/10.BC_4.IVM.P_4, MHco3/4.BC_4.IVM.P_4). Pools of male
781 and female worms were included for the parental (n = 50-60 worms) and BC_4
782 samples (n = 25-30 worms), whereas only female worms were used in the passage 3
783 & 4 samples (n = 60 worms). Genomic DNA from each pooled sample was prepared
784 using a phenol chloroform extraction protocol as previously described. Sequencing
785 libraries were prepared using a PCR-free protocol (Kozarewa et al. 2009), and
786 sequenced as described in **Table S1**. We generated approximately 614×10^9 bp of
787 sequence data, which equates to approximately 199.65× unmapped genome
788 coverage per sample. Raw sequence data quality was analysed using FASTQC
789 (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)* and was assessed using
790 MultiQC (Ewels et al., 2016) prior to further processing.

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

794 **Mapping and variant analysis**

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

801 Genome-wide variants were determined for each sample using samtools-1.3 mpileup
802 (*-F 0.25 -d 500*). Nucleotide diversity and Tajima's D was determined using *npstat*
803 (Ferretti et al., 2013), which required pileup files (which we derived from the mpileup
804 generated above) that were split per sample per chromosome as input. We
805 determined short-range linkage disequilibrium between pairs of variants in single and
806 paired reads using *LDx* (Feder et al., 2012). We compared these data with estimates
807 of LD decay over genetic distance as proposed by Hill and Weir (Feder et al., 2012;
808 Hill and Weir, 1988). *Popoolation2* (Kofler et al., 2011) was used for the analysis of
809 pairwise genetic diversity throughout the genome. Briefly, the mpileup file was
810 converted into a synchronised file (*popoolation2 mpileup2sync.jar --min-qual 20*),
811 after which indels + 5 bp (popoolation2 *identify-indel-regions.pl −−min-count 2*
812 *−−indel-window 5*) and repetitive and difficult to map regions (separately identified by
813 repeat masking the genome [http://www.repeatmasker.org/]) were excluded
814 (*popoolation2 filter-sync-by-gtf.pl*). The synchronised file was used as input to
815 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:

$$q_1' = q_1^2 + q_1 q_h + \tfrac{1}{4} q_h^2$$

$$q_h' = q_1 q_h + 2 q_1 q_0 + \tfrac{1}{2} q_h^2 + q_h q_0 \qquad [2]$$

$$q_0' = q_0^2 + \tfrac{1}{4} q_h^2$$

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

$$q_1' = 0$$

$$q_h' = q_1 + \tfrac{1}{2} q_h \qquad [3]$$

$$q_0' = q_0 + \tfrac{1}{2} q_h$$

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:

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

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

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);

911 this has been exploited to identify quantitative trait loci in yeast and malaria cross
912 populations (Abkallo et al., 2017; Parts et al., 2011). Here a heuristic approach was
913 used to identify fixed genetic differences between the parental strains before
914 modelling evolution at these loci to identify the location of alleles conveying drug
915 resistance.

916 Three filters were used to identify putatively fixed differences between parental
917 strains. Firstly, we identified genomic sites for which the frequency of the resistant
918 allele was 95% or greater in the resistant parent, requiring a read depth of at least
919 50x for such sites to achieve a good level of statistical certainty. Secondly, we noted
920 that, following the backcross performed at the start of the experiment, no locus can
921 be homozygous for the resistant allele and as such, the resistant allele frequency
922 can be no greater than 50% in the population; by way of reducing noise, we removed
923 any locus with resistant alleles at a frequency of 60% or greater in the
924 MHco3/10.BC_4.noIVM sequence data. Thirdly, we noted that at homozygous
925 separating sites in a population following a cross, the allele frequency will change
926 smoothly over time; where there are N genomes in the population, the allele
927 frequency, considered as a function of chromosome position, will change by $1/N$
928 wherever a crossover recombination event occurs within a genome. A diffusion
929 model of allele frequency change, described in a previous publication (Abkallo et al.,
930 2017), was used to identify allele frequencies across the genome that were
931 consistent with this pattern. This analysis fits a posterior distribution to the allele
932 frequencies across the genome, also inferring the extent of noise in the sequence
933 data via a beta-binomial model intrinsic to the fitting process. Conservatively, sites no

934 more than eight standard deviations from the mean of the posterior frequencies were
935 included in the analysis. These approaches resulted in data from a total of 1,368 loci
936 in the MHco10(CAVR) dataset, and from 219 loci in the MHco4(WRS) dataset, were
937 retained; the full set of allele frequencies are shown in Supporting Figure S5.

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

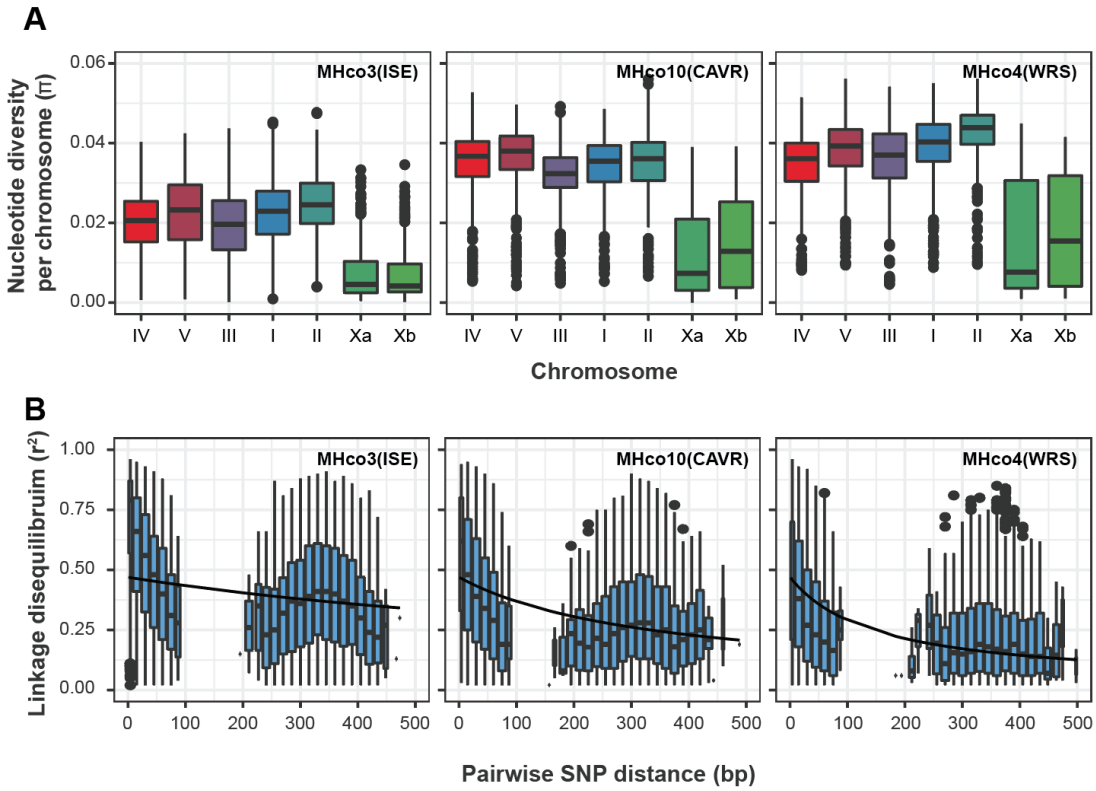
947 $$w_1 = 1 + s$$
948 $$w_h = 1 + hs \qquad [10]$$
949 $$w_0 = 1$$

950 where s is the selection coefficient (fitness advantage of the homozygous resistant
951 compared to homozygous susceptible) and h is the dominance coefficient
952 (proportion of fitness change conveyed by a single copy of the allele). We assumed
953 that all worms have equal fitness in the absence of the drug. The likelihood of any
954 given inference was calculated using the beta-binomial likelihood function obtained

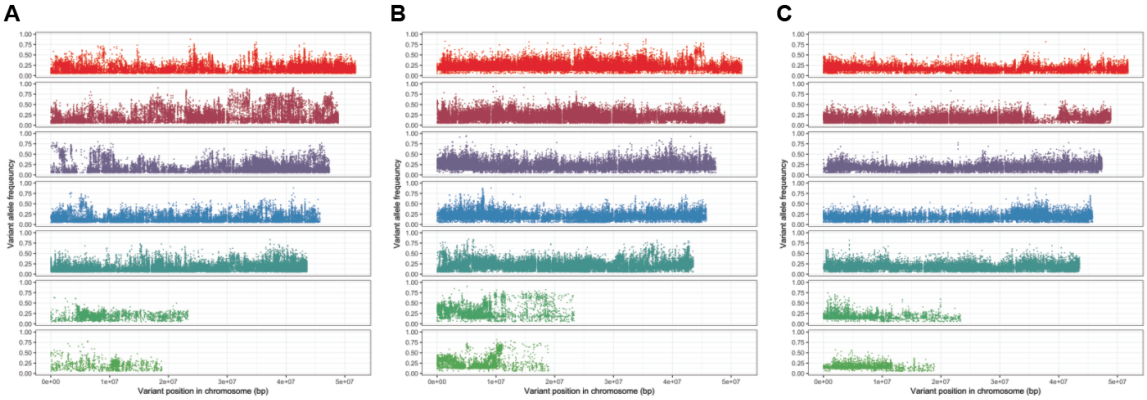
from the diffusion model above. Accounting for the stochasticity of the system, a set of at least 250 simulations were generated for each set of parameters, calculating the mean likelihood fit to the data across the simulations. A bootstrapping process was applied to quantify the uncertainty in this 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. 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.

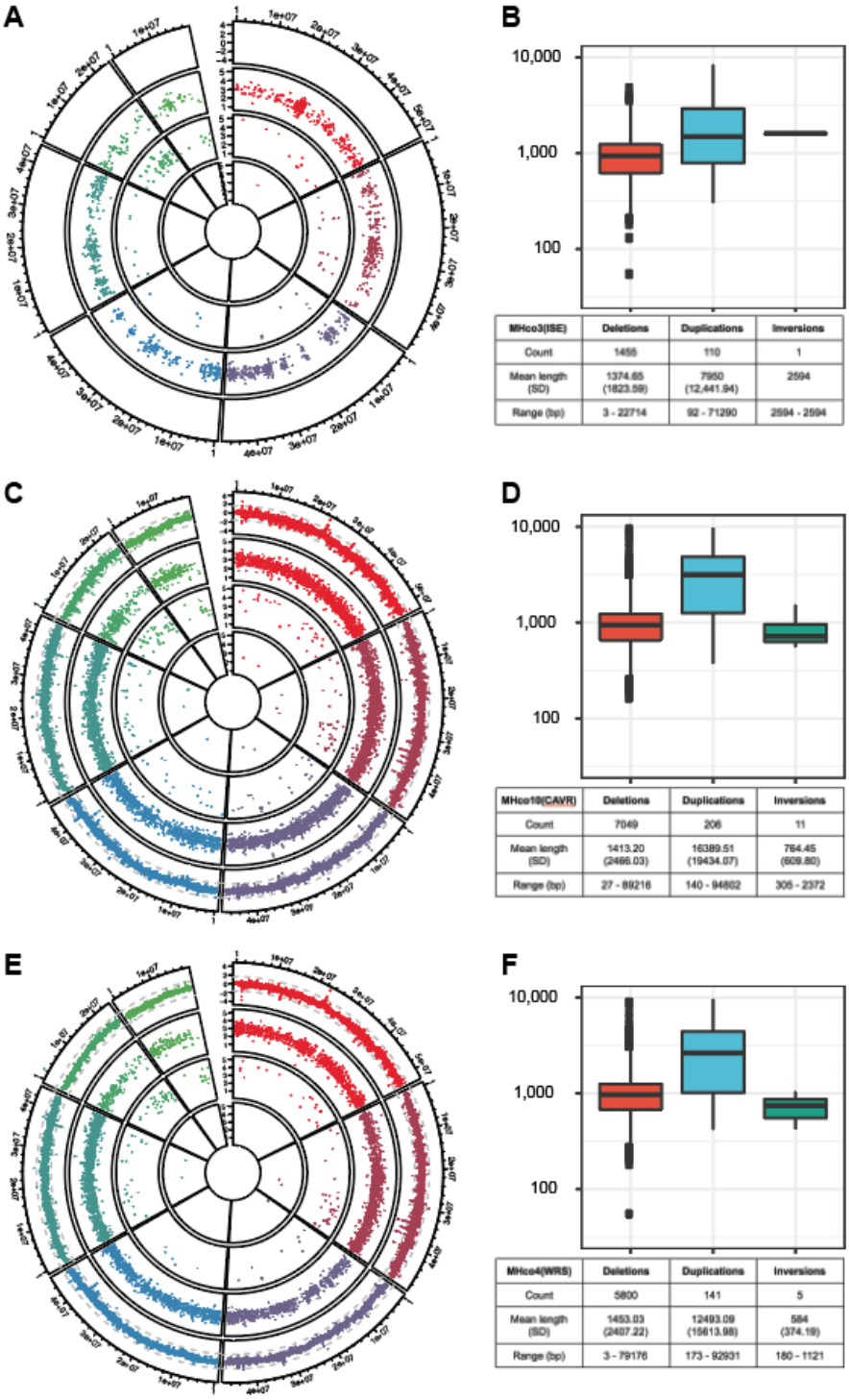
964 **Table S1. Sample sequencing data archived at European Nucleotide Archive**
965 **repository under the study accession PRJEB2353**
966 - See accompanying .xlsx document



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



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

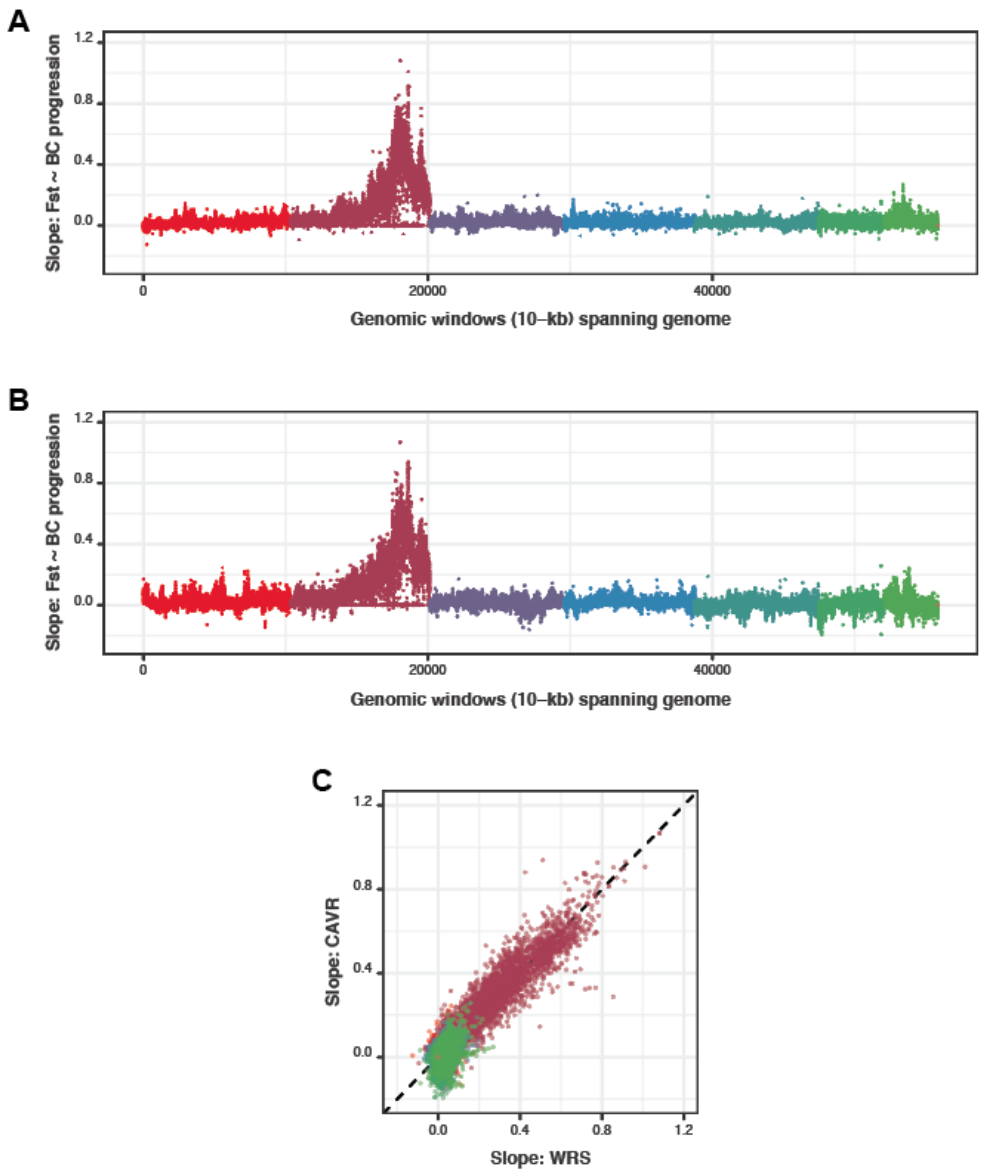


MHco3(ISE)	Deletions	Duplications	Inversions
Count	1455	110	1
Mean length (SD)	1374.65 (1823.59)	7950 (12,441.94)	2594
Range (bp)	3 - 22714	92 - 71290	2594 - 2594

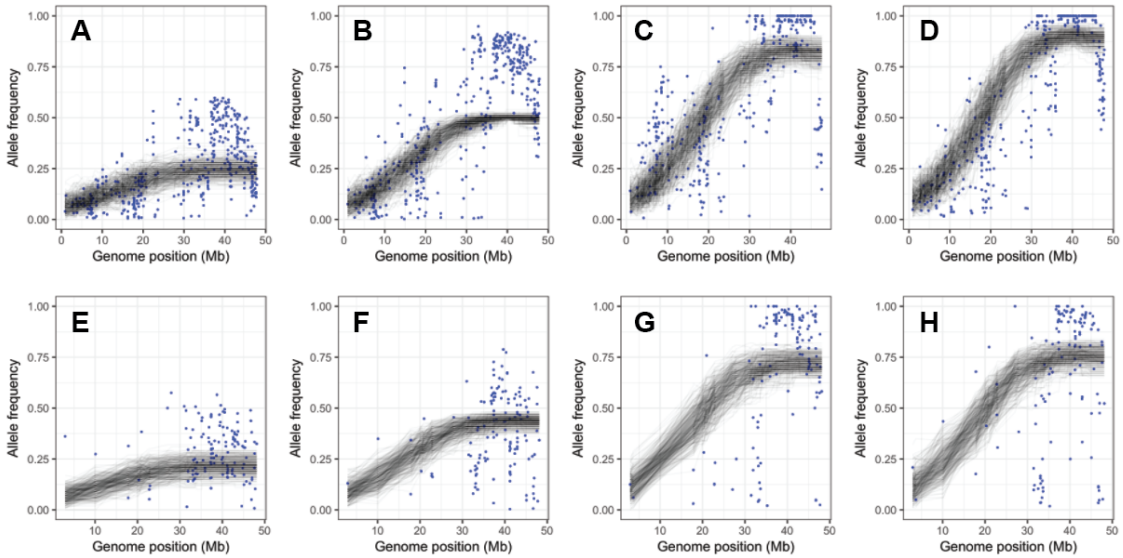
MHco10(CAVR)	Deletions	Duplications	Inversions
Count	7049	206	11
Mean length (SD)	1413.20 (2466.03)	16389.51 (19434.07)	764.45 (609.80)
Range (bp)	27 - 89216	140 - 94802	305 - 2372

MHco4(WRS)	Deletions	Duplications	Inversions
Count	5800	141	5
Mean length (SD)	1453.03 (2487.22)	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**.



983 **Figure S4. Summary of genome-wide change in F_{ST} throughout the backcross**
984 **and subsequent passage.** A regression between F_{ST} and the four sampling time
985 points was performed for each 10 kb window sampled across the genome for both
986 MHco3/10 (**A**) and MHco3/4 (**B**) lines. The slope of the regression was plotted. **C**.
987 Correlation in slope (F_{ST} vs backcross progression) between MHco3/10 and
988 MHco3/4. Dashed line represents x=y. Colours represent chromosomes as
989 described in Figure 2A.



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

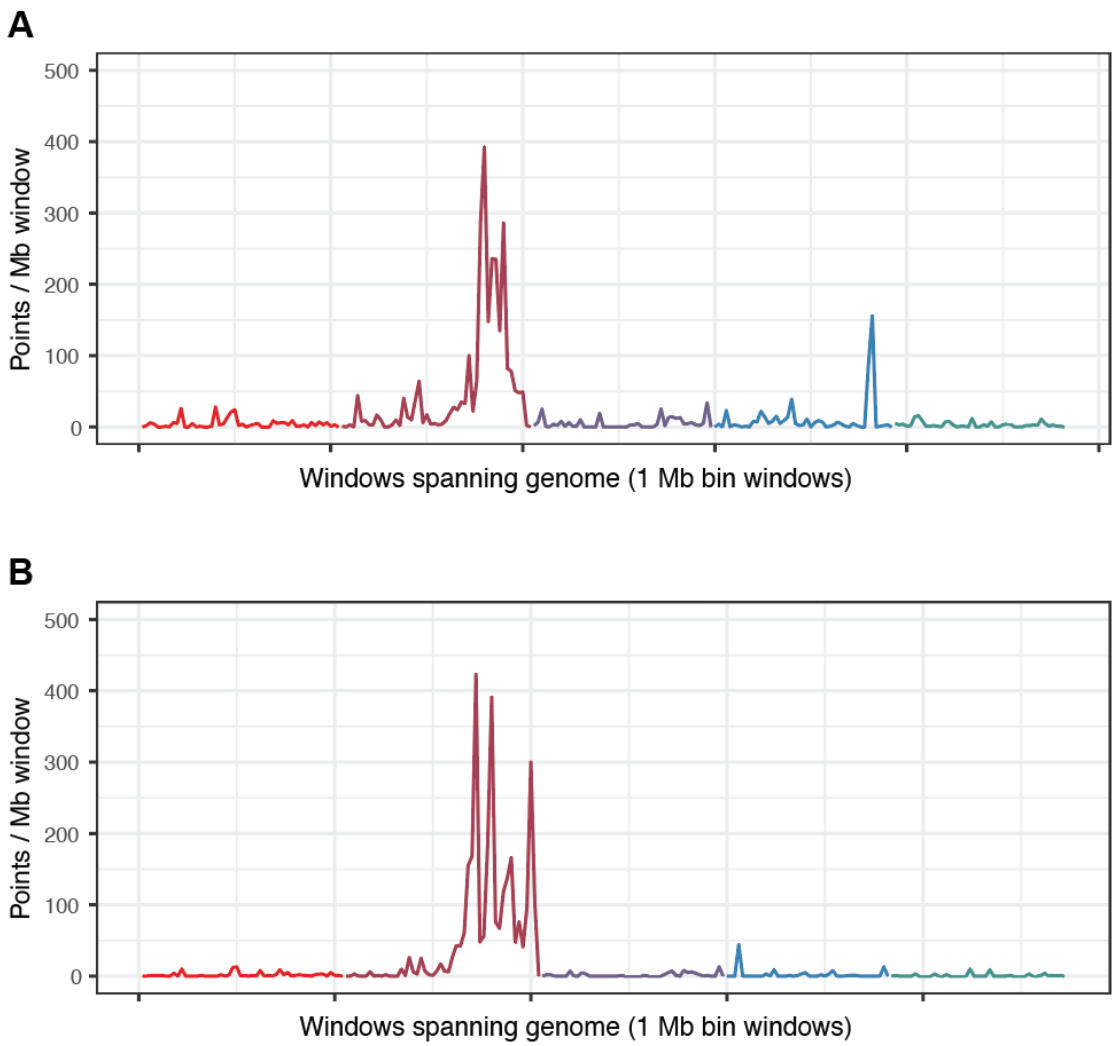


Figure S6. Location of significantly non-neutral loci identified using the single-locus population genetic model. Corresponding peaks in the location of significant sites can be seen in the MHco3/10 (**A**) and MHco3/4 datasets (**B**). A total of 70.6% of significant sites in the MHco3/10 dataset, and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

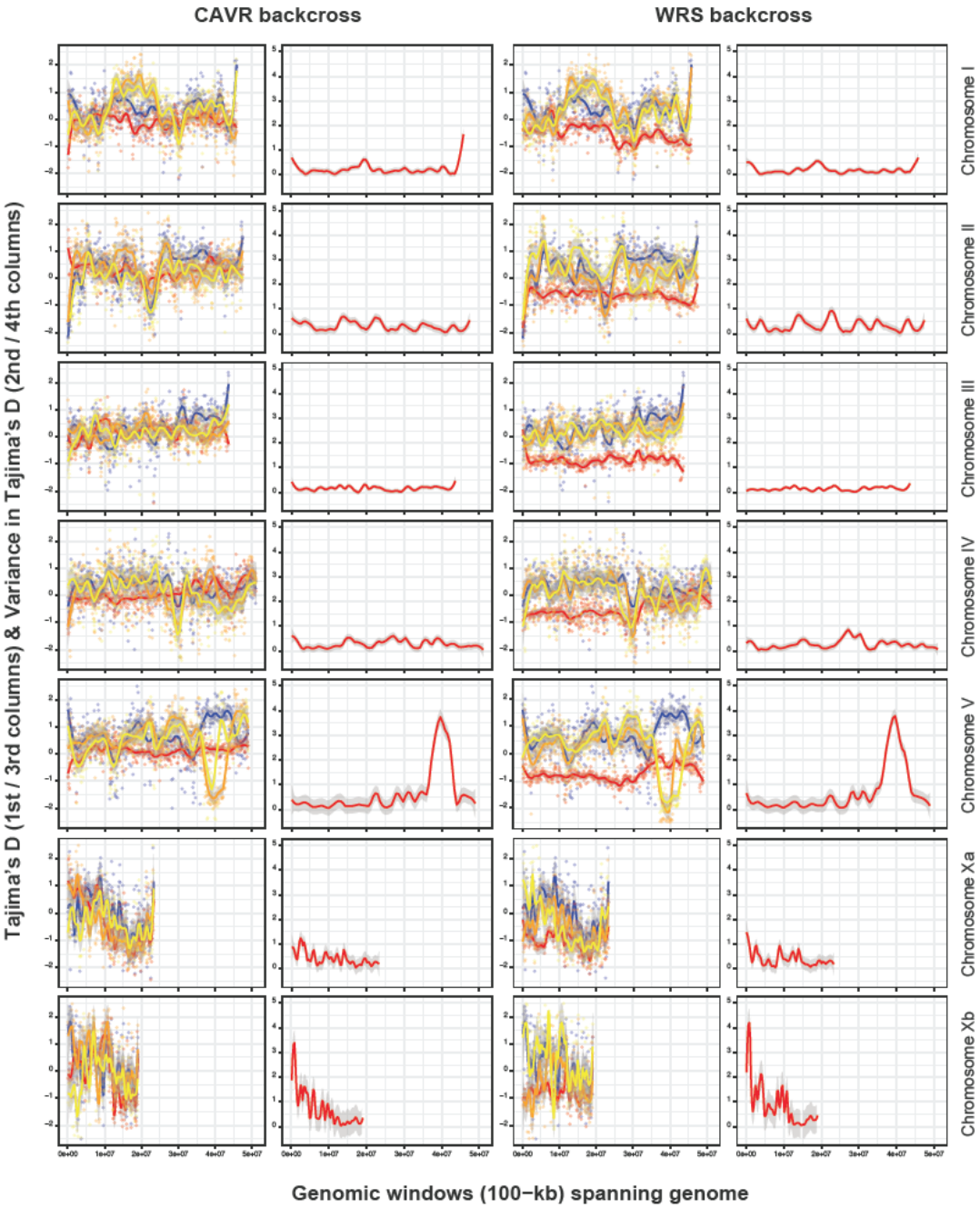
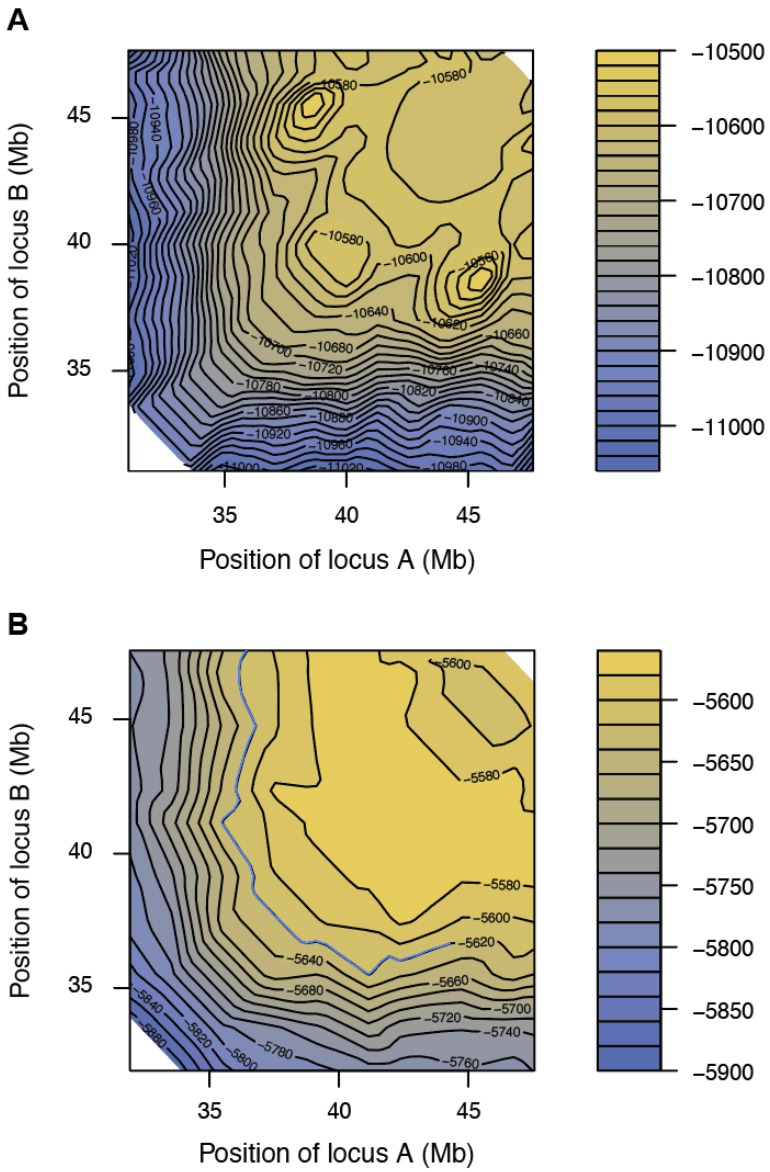
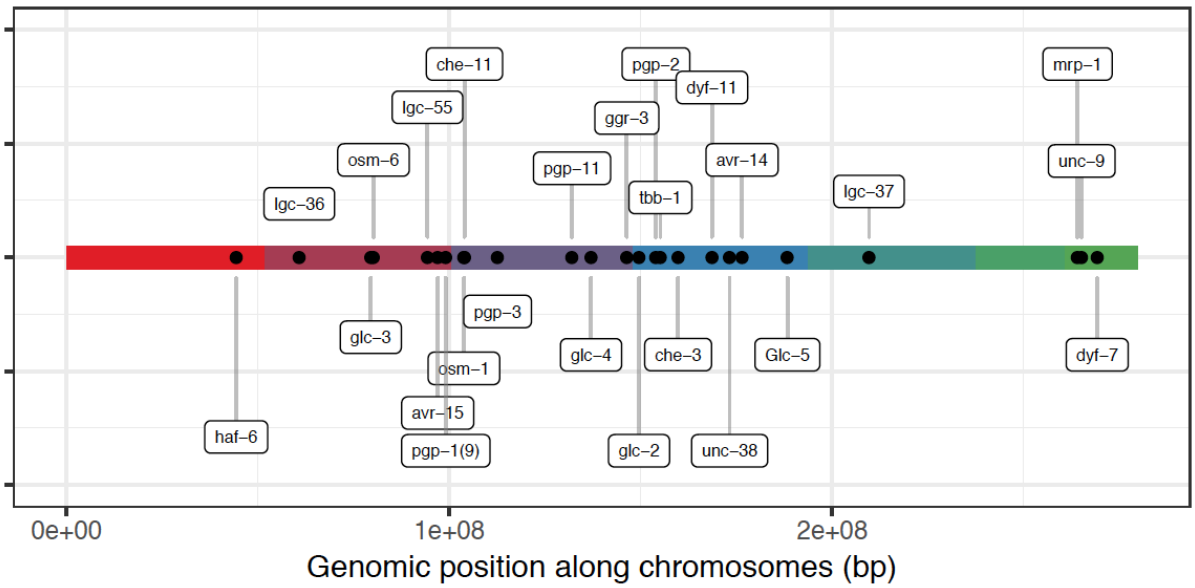


Figure S7. Analysis of Tajima's D variation in each chromosome per cross. Comparison of Tajima's D per chromosome between MHco3(ISE) parent (blue), MHco10(CAVR) (panel column 1; red) or MHco4(WRS) (panel column 3; red) and passages 3 (orange) and 4 (yellow) of the crosses. Tajima's D was calculated using *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.



1011 **Figure S8.** Contour maps of log likelihood scores derived from the two locus driver
1012 model. **A.** MHco10(CAVR). **B.** MHco4(WRS). The model was restricted to
1013 interactions between pairs of loci at least 2 Mb apart.



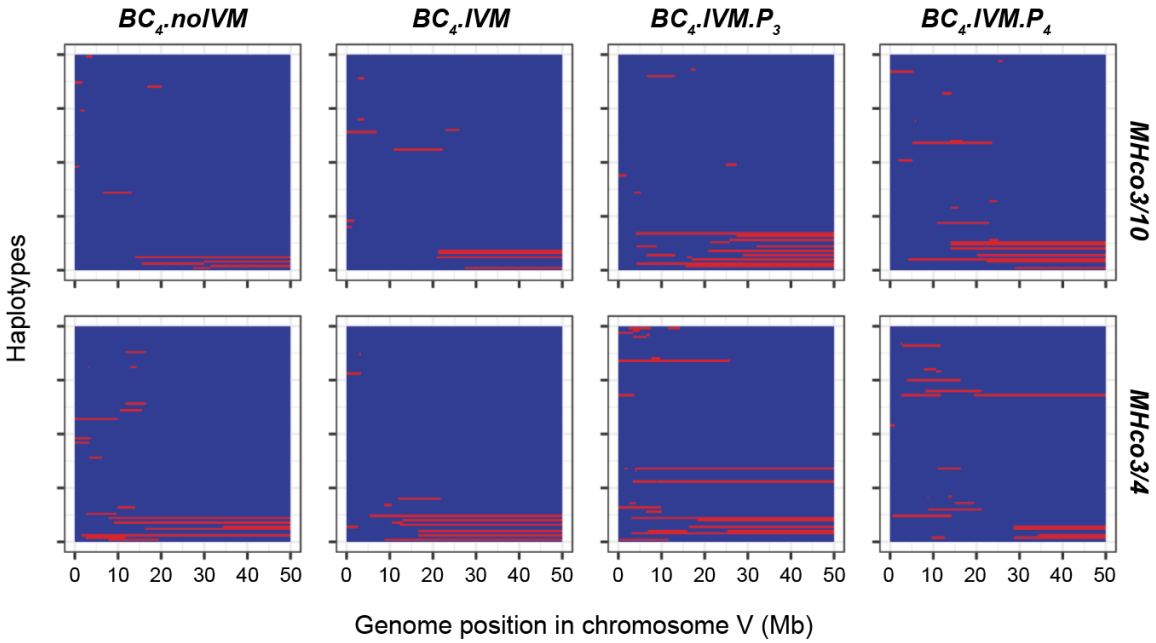
1014 **Figure S9. Relative position and location of candidate genes from the literature**
1015 **proposed to be associated with ivermectin resistance in *Haemonchus***
1016 ***contortus* and / or *Caenorhabditis elegans*.** Gene coordinates are presented in
1017 Table S2. Colours represent chromosomes as described in Figure 2A.

1018 **Table S2. Candidate genes from literature proposed to be associated with**
1019 **ivermectin resistance in *Haemonchus contortus* and / or *Caenorhabditis***
1020 ***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.



1222 **Figure S10: Haplotype structure of chromosome V in an example output from**
1223 **the model under neutral evolution.** Segments of genome from the resistant parent
1224 are shown in red, while segments of genome from the susceptible parent are shown
1225 in blue. The repeated backcross removes most of the resistant genotypes from the
1226 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.