Genomic signatures of selection associated with benzimidazole drug treatments in *Haemonchus contortus* field populations

Janneke Wit[1*], Matthew L. Workentine[2], Elizabeth Redman[1], Roz Laing[3], Lewis Stevens[4], James A. Cotton[5], Umer Chaudhry[6 #], Qasim Ali[7], Erik C. Andersen[8], Samuel Yeaman[9], James D. Wasmuth[10], John S. Gilleard[1*]

[1] Department of Comparative Biology and Experimental Medicine, Host -Parasite Interactions (HPI) program, University of Calgary, Calgary, Alberta, Canada

[2] Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

[3] Institute of Biodiversity Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Garscube Campus, Glasgow, UK

[4] Tree of Life, Wellcome Sanger Institute, Cambridge, UK

[5] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK. CB10 1SA.

[6] University of Edinburgh, Roslin Institute, Easter Bush Veterinary Centre, Roslin, Midlothian, UK

[7] Department of Veterinary Parasitology, University of Veterinary and Animal Sciences Lahore, Pakistan

[8] Molecular Biosciences, Northwestern University, Evanston, IL, USA

[9] Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada

[10] Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

[#] Present address: School of Veterinary Medicine, University of Surrey, England, UK

*Corresponding authors: jannekewit@gmail.com, jsgillea@ucalgary.ca

ABSTRACT

Genome-wide methods offer a powerful approach to detect signatures of drug selection in parasite populations in the field. However, their application to parasitic nematodes has been limited because of both a lack of suitable reference genomes and the difficulty of obtaining field populations with sufficiently well-defined drug selection histories. Consequently, there is little information on the genomic signatures of drug selection for parasitic nematodes in the field and on how best to detect them. This study was designed to address these knowledge gaps using field populations of *Haemonchus contortus* with well-defined and contrasting benzimidazole-selection histories, leveraging a recently completed chromosomal-scale reference genome assembly. We generated a panel of 49,393 ddRADseq markers and used this resource to genotype 20 individual *H. contortus* adult worms from each of four *H. contortus* populations: two from closed sheep flocks that had an approximately 20-year history of frequent treatment exclusively with benzimidazole drugs, and two populations with a history of little or no drug treatment. The populations were chosen from the same geographical region to limit population structure in order to maximize the sensitivity of the approach. A clear signature of selection was detected on the left arm of chromosome I centered on the isotype-1 β-tubulin gene in the benzimidazole-selected but not the unselected populations. Two additional, but weaker, signatures of selection were detected; one near the middle of chromosome I and one near the isotype-2 β-tubulin locus on chromosome II. We examined genetic differentiation between populations, and nucleotide diversity and linkage disequilibrium within populations to define these two additional regions as encompassing five genes and a single gene. We also compared the relative power of using pooled versus individual worm sequence data to detect genomic selection signatures and

how sensitivity is impacted by sequencing depth, worm number, and population structure.

In summary, this study used *H. contortus* field populations with well-defined drug selection histories to provide the first direct genome-wide evidence for any parasitic nematode that the isotype-1 β-tubulin gene is the quantitatively most important benzimidazole resistance locus. It also identified two additional genomic regions that likely contain benzimidazole-resistance loci of secondary importance. Finally, this study provides an experimental framework to maximize the power of genome-wide approaches to detect signatures of selection driven by anthelmintic drug treatments in field populations of parasitic nematodes.

AUTHOR SUMMARY

Benzimidazoles are important anthelmintic drugs for human and animal parasitic nematode control with ~0.5 billion children at risk of infection treated annually worldwide. Drug resistance is common in livestock parasites and a growing concern in humans. *Haemonchus contortus* is the most important model parasite system used to study anthelmintic resistance and a significant livestock pathogen. It is also one of the few parasitic nematodes with a chromosomal-scale genome assembly. We have undertaken genome-wide scans using a dense RADseq marker panel on worms from natural field populations under differing levels of benzimidazole selection. We show that there is a single predominant genomic signature of selection in *H. contortus* associated with benzimidazole selection centred on the isotype-1 β-tubulin locus. We also identify two weaker signatures of selection indicative of secondary drug resistance loci. Additionally, we assess the minimum data requirements for parameters including worm number, sequence depth, marker density needed to detect the signatures of selection and compare individual to Poolseq analysis. This work is the first genome-wide study in a parasitic nematode to provide direct evidence of the isotype-1 β-tubulin locus being the single predominant benzimidazole resistance locus and provides an experimental framework for future population genomic studies on anthelmintic resistance.

1. INTRODUCTION

Parasitic nematode infections are of major medical and agricultural importance worldwide [1,2]. Control strategies largely depend on the regular use of anthelmintic drugs, which has caused widespread anthelmintic drug resistance in many parasitic nematode species of livestock [1,3]. Additionally, concerns are increasing that mass drug administration programs are selecting for anthelmintic resistance in human helminths [4–8]. Over the last 30 years, there has been a large amount of research into the molecular genetic basis of anthelmintic resistance, particularly in gastrointestinal nematodes of livestock [9]. Most of this work has been dominated by candidate gene studies, where certain genes are prioritized based on knowledge of the drug mode-of-action [10–14] or by the extrapolation of genetic studies in the model organism *Caenorhabditis elegans* [15,16]. Whilst this approach has been successful in some cases, most notably for the benzimidazole drug class, it has been much less successful in others [17]. For example, in the case of the macrocyclic lactones, which is one of the most important broad spectrum classes of anthelmintic drugs, the evidence implicating the leading candidate genes has been inconsistent between studies [18–22].

Genome-wide approaches, which make no *a priori* assumptions about the underlying mechanisms of resistance, are potentially more powerful approaches to identify anthelmintic resistance loci than candidate gene studies. The most successful example in a helminth to date was the use of classical linkage mapping to identify a Quantitative Trait Locus (QTL) for oxamniquine (OXA) resistance in the human trematode *Schistosoma mansoni* [23]. Subsequent work confirmed the functional importance of a mutation in a sulfotransferase enzyme (SmSULT-OR) gene to the OXA-resistant phenotype [23,24].

The small ruminant parasite *Haemonchus contortus* is the leading parasitic nematode model for anthelmintic resistance research for a variety of reasons, including the availability of well characterized anthelmintic resistant isolates, a good understanding of its genetics, and the ability to undertake genetic crosses [9,25–27]. A chromosomal-scale reference genome assembly and annotation enables genome-wide approaches in this parasitic nematode species (Doyle et al 2020). This new reference genome was successfully used to map the major ivermectin resistance QTL on chromosome V in two independent laboratory passaged *H. contortus* strains by a serial backcrossing approach [28,29]. However, such genetic crosses are extremely challenging in parasitic nematodes compared to model organisms and have only been demonstrated in a few parasite species [26,30]. An alternative way to map the genetic loci underlying anthelmintic resistance is to use population genomic approaches on field populations. This approach is not only technically simpler and more scalable but is also more likely to identify anthelmintic resistance loci that are relevant to field populations.

As reference genomes for parasitic nematodes continue to improve, and DNA sequencing costs reduce, population genomic studies on natural field populations should become increasingly feasible. However, there is currently a lack of data from field populations with sufficiently well-defined and contrasting drug treatment histories to unambiguously link signals of selection in the genome with specific drug treatments. For example, most *H. contortus* field populations have been subject to treatment with multiple different drug classes and to significant animal movement. Consequently, there is a need for studies on field populations with well-defined drug treatment histories to elucidate the genomic location and nature of signatures of selection associated with specific drug classes.

It is now well established that mutations in the isotype-1 β-tubulin gene are important determinants of benzimidazole resistance in *H. contortus* and other strongylid nematodes [10,31–36]. However, a key knowledge gap at present is whether the isotype-1 β-tubulin locus is the only major locus under selection by benzimidazole drug treatments or whether there are additional selected loci and, if so, their relative importance [37]. Interestingly, a recent study in which whole-genome sequencing was applied to 233 individual adult *H. contortus* worms from field populations, with varied and complex drug selection histories, identified multiple regions of the genome that showed signatures of selection including around the isotype-1 β-tubulin locus [38]. However, this study could not associate these signatures of selection with treatment by a specific drug class as the populations used had a variety of complex drug treatment histories. A major aim of our current study was to determine which loci in the *H. contortus* genome show strong evidence of selection that are specifically associated with benzimidazole selection in the field.

Another aim of this study was to take advantage of the clear and contrasting benzimidazole drug selection histories of the *H. contortus* populations on government and rural farms in Pakistan to investigate the best technical approach to detect genomic signatures of drug selection in field populations. Reduced-representation sequencing methods, where a random but consistent subset of the genome is sequenced, are potentially powerful and cost-effective for genome-wide methods [39]. Larger numbers of organisms and/or samples can be sequenced than by whole-genome sequencing at the same cost, and while sequencing costs are dropping, cost is an important consideration when examining larger sample sets, particularly for laboratories outside of the major genome sequencing centres [40,41]. One of the more widely used reduced-representation approaches is Restriction-site Associated DNA sequencing (RADseq) [42–44]. This

150 approach has been extremely successful at investigating population diversity and
151 differentiation in a wide range of organisms [45–48] but it is still a challenge to get a high-
152 density marker set to identify signatures of selection [49–51].

153 Here, we describe the development of a high density RADseq marker-set, and its
154 use to investigate the genomic signature(s) of benzimidazole drug selection in *H.*
155 *contortus* field populations. The strongest signature of benzimidazole drug selection, by
156 far, that was present in two populations with a history of regular benzimidazole
157 treatments, but absent in two untreated populations, was in the genomic region
158 surrounding the isotype-1 β-tubulin gene. Two other genomic regions under drug
159 selection were identified but the signals of selection at these loci were weaker. In
160 addition, we investigated the effect of sample number and read depth, and compared the
161 analysis of individual worm sequence data with pooled sequence data on the ability of
162 RADseq to detect genomic selection signatures in order to provide a framework for future
163 experimental design.

164

165 **2. MATERIALS AND METHODS**

166 **2.1 Parasite material**

167 Populations of adult *H. contortus* worms were harvested from four different sheep
168 flocks in the Punjab province of Pakistan [52]. Two populations were sampled from
169 government farm flocks with a history of regular BZ treatment over several decades;
170 Treated 1 (T1) from Okara and Treated 2 (T2) from Jahangirabad (T1 is Pop1S and T2 is
171 Pop3S in [52]). These government farms established their sheep flocks using local sheep
172 breeds (lohi and kajli) in 1985 and 1989 respectively and have been alternatively treated
173 with albendazole and oxfendazole, but not other anthelmintic classes, approximately
174 every three months since their establishment. These herds have been closed to animal

175 movement since they were established. There has been some historical movement of
176 stock between the farms, but no animals have been introduced from elsewhere. Two
177 other populations were sampled from rural flocks with no history of drug treatment, from
178 two abattoirs in the same region; Untreated 1 (U1) and Untreated 2 (U2). Worms were
179 harvested directly from ewe abomasa at necropsy at abattoirs from sheep carcases
180 following routine slaughter for human consumption and stored in 70% ethanol at -80˚C.
181 For this study, 20 females were randomly selected from each sample.
182
183 **2.2 DNA extraction**
184 Adult female worms were rehydrated in sterile water for 60 minutes prior to
185 extraction. Tissue was isolated from the anterior portion of the adult worm, avoiding the
186 gonads containing progeny, and DNA was extracted using a protocol adapted from Bennet
187 *et al.* [53]. After dissection of the animal, the anterior portion was transferred to 600 µl
188 lysis buffer (384 µl H2O, 60 µl 1M Tris HCl (pH 8.5), 60 µl 1% SDS, 60 µL 0.5M EDTA (pH
189 7.5), 12 µl 5M NaCl, 12 µl proteinase K (14-22 mg/mL, Qiagen), and 12 µl β–
190 mercaptoethanol). Samples were incubated overnight at 55˚C before adding 2 µl of
191 RNaseA (100 mg/ml, Qiagen) and incubated for 15 min at 45˚C. After adding 150 µl 5M
192 NaCl and 100 µl CTAB/NaCl samples were incubated at 65˚C for 15 min. An equal volume
193 of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) was added and the solution was
194 incubated for 1.5 hr on a slow rocker at room temperature (RT). Samples were
195 centrifuged at 3000 g for 15 min at 12˚C and the aqueous phase was collected. An equal
196 volume of chloroform:isoamyl alcohol (24:1, Sigma) was added to the aqueous phase and
197 the solution incubated for 1 hour at RT, centrifuged as before and the aqueous phase
198 collected. Two volumes of ice-cold ethanol were added and gently mixed. The solution
199 was placed at -20˚C overnight before being centrifuged as before at 4˚C. After removal of

the supernatant, the pellet was washed twice with 70% EtOH. After the final wash, the sample was centrifuged at 14000 rpm for 10 min, the supernatant removed, centrifuged for an additional minute to remove the remainder of the ethanol and the resulting pellet air dried before resuspension in 30 µl water and incubation overnight at 4˚C. Sample quality was checked by quantification on the Qubit system (ThermoFisher) and A_{260}/A_{280} absorbance ratios on a NanoVue (Biochrom Spectrophotometers). 2.5 µl per sample (46.4-400 ng) was added to a REPLI-g Single Cell Kit (Qiagen) for whole genome amplification (WGA) following the manufacturers protocol for Amplification of Purified Genomic DNA.

2.3 Amplicon sequencing of isotype-1 β-tubulin and isotype-2 β-tubulin genes and haplotype network analysis

Amplicon sequencing was used to determine the frequency of single nucleotide polymorphisms (SNPs) in the isotype-1 and isotype-2 β-tubulin genes previously associated with benzimidazole resistance in *H. contortus* - F200Y (TTC>T**A**C), F167Y (TTC>T**A**C), E198A (GAA>G**C**A), or E198L (GAA>**TT**A) – in each population. The amplicon sequencing protocol and data analysis has been previously described in Avramenko et al. [54,55] and adapted for isotype-2 -tubulin with primers in **File S1**. For the haplotype analysis of isotype-1 β-tubulin, reads classified as *H. contortus* by the Mothur [56] pipeline in the amplicon data analysis were extracted from the original FASTQ files. Primers were removed with Cutadapt [57], and haplotype analysis was conducted with DADA2 [58,59], following their pipeline tutorial for demultiplexed Illumina paired-end reads. Resulting amplicon sequence variants (ASVs) were used to create a haplotype network using ape and pegas in R [60,61].

2.4 Preparation and sequencing of double-digest single worm RAD libraries

The single worm WGA samples were prepared for sequencing following a protocol adapted from Peterson *et al.* (2012). Post-WGA samples were purified using a standard bead clean-up protocol using AMPure XP beads (Beckman), with 1.5x beads instead of 1.8x sample:bead volume, and eluted in water. The DNA was digested with *Mlu*CI (5µl Cut smart buffer, 1.5-2 µg DNA, 3 µl *Mlu*CI (New England Biolabs Inc. (NEB), and water to make up 49 µl) for three hours at 37˚C after which 3 µl *Nla*II (NEB) was added and the reaction incubated at 37˚C for a further three hours. The samples were cleaned as before. Barcoded adapters were ligated in a 50 µl reaction volume, with 4 µl T4 buffer, 7.5 µl each of adapter P1 and P2, 0.7 µl T4 ligase (2 M U/ml, NEB), 250 ng DNA (adapters in **File S1**) and water. Libraries from individual worms of different populations were randomized to create pools containing 20 individuals each for sequencing, cleaned as before, and size-selected on a PippinPrep (Sage Science) using the standard protocol for Tight Range (190 bp) on a 3% internal standards cartridge, resulting in a fragment range of 190-300 bp with a peak at 210 bp. Quality and fragment size was checked on a Tape Station (Agilent), and P5 and P7 adapters were ligated per pool following the KAPA HiFi Hotstart PCR kit standard protocol (Kapa Biosystems). Pooled samples were then combined into one sample and cleaned as before. Because the fragments ranged in size, quantity was checked in triplicate using a Qubit broad range kit, rather than a qPCR approach, and diluted to 4 nM after correction for average size.

The pooled individual worm libraries were sequenced on the Illumina NextSeq platform. Populations U1, U2, T1, and T2 were sequenced on one High Output 2 x 75 bp kit. The U2 population was discarded after only 20.22 - 53.54 % of its reads were mapped to the reference with the remainder being bacterial contamination. Consequently, the U2

249 population was re-sequenced using a Mid Output 2 x 75 bp kit. The loading concentration
250 was 1.2 pM, and 50% PhiX was spiked in for diversity due to the low diversity library.
251 Fragment-size selection was based on predicted fragment recovery with the draft
252 genome assembly (GCA_000469685.1 [62]) using SimRAD [63].
253
254 **2.5 Quality Control and Analysis of Sequence Data**
255 The data were demultiplexed using process_radtags in the Stacks pipeline [64]
256 (SRA accession number PRJN822658). All samples were assessed for general sequence
257 quality using FastQC [65], before removing adapters using Cutadapt [57]. Read quality
258 was high across the full read length for both sequencing runs; therefore, no hard
259 trimming was conducted. Reads were aligned against the chromosomal-scale *H. contortus*
260 reference genome assembly (GCA_000469685.2, N50 = 47.38 Mb, N50(n) = 3, n= 7 [66])
261 using Bowtie2 (--local --very-sensitive-local -X 300 -I 80 [67]). A BAM file was generated
262 for each sample using Samtools by filtering for paired reads only and then sorting [68].
263 The sorted data was analyzed using gstacks and populations using the –fstats flag from
264 the Stacks pipeline. Only SNPs that were present in three of four populations, in 50% of
265 individuals per population, and at a minimum frequency of 5% were retained in the
266 output. In addition, only the first SNP per RAD locus was reported to avoid linkage and
267 variable SNP count within a locus to affect the interpretation of the data.
268 To assess the robustness of the RADseq analysis and investigate the limits to its
269 ability to detect signatures of selection, additional analyses were conducted using the
270 same basic parameters but fewer individual worms per population (5, 10, and 15,
271 randomly selected), as well as fewer reads per individual (1M, 1.5M, and 2M reads where
272 possible, randomly selected using seqtk sample). Finally, the raw reads were realigned to
273 an earlier version of the reference genome (GCA_000469685.1, N50 = 0.099 Mb, N50(n)

= 876, n= 12,915 [62]), to illustrate the effect of genome quality on identifying regions of interest in the genome.

2.6 Population differentiation, expected heterozygosity, and linkage disequilibrium analyses to detect signatures of selection

All analyses were performed on data generated by Stacks (version 2.2), reporting one SNP per RAD-locus, unless otherwise stated. To examine overall population differentiation, population level F_{ST} values were calculated and a Principal Component Analysis was conducted on all reported SNPs across the populations using the vcfR (version 1.8.0) and adegenet (version 2.1.1) packages in R [69,70].

To detect regions in the genome where population differentiation was significantly higher than background noise, we used a top candidate method as described in Yeaman et al. [71]. The 99th percentile of all F_{ST} values was determined and F_{ST} values were then binned over 150 kb intervals. Top candidate regions were identified as bins with an exceptional proportion of their total SNPs being F_{ST} outliers based on a binomial distribution. For expected heterozygosity, a similar approach was undertaken. Outliers were defined as expected heterozygosity values falling in the lowest 25th percentile and then binned across 150 kb regions. Top candidate bins were identified as those having a higher than expected proportion of low expected heterozygosity based on a binomial distribution. Linkage disequilibrium per population was calculated in R using the cor() function with "pairwise.complete.obs". These values were binned across 150 kb regions and the percentage of R^2 values above the 99th percentile of all R^2's per population was recorded.

297 **2.7 Analysis of mock pooled sequence data**

298 All RADseq data from individual worms per population were pooled to create
299 mock pooled sequence data. Data were pooled after demultiplexing with process_radtags,
300 as these RAD-specific barcodes would not be present in Illumina sequence data of pooled
301 samples. Samples were then aligned with Bowtie2 as above. The resulting aligned files
302 were subsampled randomly using seqtk, resulting in 100× genome coverage. For
303 analyses of individual samples, average expected heterozygosity was determined for one
304 SNP per RAD-locus, therefore bedtools was used to concatenate SNPs within 150 bp of
305 each other to one marker for pooled samples [72,73]. These subsamples were analyzed
306 using PoPoolation Variance-sliding.pl (Variance-sliding.pl --measure pi --min-count 2 --
307 min-qual 20 --min-coverage 15 --max-coverage 400 --pool-size 20 --window-size 1 --
308 step-size 1) and Popoolation2 snp-frequency-diff.pl (snp-frequency-diff.pl --min-count 3
309 --min-coverage 15 --max-coverage 400 [74,75], generating Tajima's Pi and F_{ST} statistics.
310

311 **2.8 Detecting genes in genomic regions under selection**

312 All genes and corresponding proteins within the genomic regions showing
313 consistent signatures of selection in the treated populations across multiple estimates
314 were extracted from the gene annotation file
315 haemonchus_contortus.PRJEB506.WBPS14.annotations.gff3 and the FASTA file
316 haemonchus_contortus.PRJEB506.WBPS14.protein.fa. Potential causal variation in these
317 genes would not necessarily be identified in RAD-data as a result of limited sequence
318 coverage of the region. Therefore functional annotation of the selected genes was
319 performed with InterPro version 5.35-74.0 and Pfam database version 32.0
320 (interproscan.sh -i [input_protein.fa] -d out -dp -t p --goterms -appl Pfam -f tsv).
321

322 **2.9 Data availability.**

323 All samples for the ddRADseq project are collected under the BioProject accession
324 number PRJNA822658.

325

326 **3. RESULTS**

327

328 **3.1 Developing a panel of genome-wide RADseq markers for *H. contortus***

329 Genome-wide RADseq markers were generated by digestion of the genome with
330 *MluC*I and *Nla*II, fragment size selection of 190-300 bp (peak at 210 bp), and sequencing
331 on an Illumina NextSeq (2x75bp). On average, 77.7% of the reads mapped to the
332 reference genome sequence (max: 81.4%, min: 60.2%, median: 78.6%). Of the mapped
333 reads, 77.6% passed the paired-read filter providing an overall average total of 60.33%
334 of raw reads for further analysis using Stacks [64] number of reads kept: average =
335 3,337,301, median = 3,186,605, min = 1,283,698, max = 5,501,942). We kept only those
336 markers present in 50% of individuals per population and in all three populations (75.4%
337 of all primary alignments). Average coverage and marker density for all comparisons are
338 shown in **Table 1** and **Fig. S1**. With baseline parameters, the range of reads per individual
339 was 4.4X- 14.2X coverage.

340

Dataset	**Average coverage per marker**	**Marker count**
Mapping to reference genome: GCA_000469685.2		
Baseline *(50% individuals, 3 populations)*	9.1	49,393
Stringent *(75% individuals, 3 populations)*	9.1	26,269
Mapping to reference genome: GCA_000469685.1		
Baseline *(50% individuals, 3 populations)*	9.8	42,987
Data Subsets *(50% individuals, 3 populations)*		
Population size = 15	9.0	45,226
Population size = 10	9.4	49,687
Population size = 5	9.0	52,532
Read depth = 2.0M/sample	7.4	42,818
Read depth = 1.5M/sample	6.3	37,432
Read depth = 1.0M/sample	5.0	30,136
Read depth = 0.5M/sample	3.3	18,283

Read depth = 0.25M/sample	2.3	7,769
Read depth = 0.1M/sample	1.6	1,666

Table 1. Overview of coverage and average marker recovery across different analysis parameters and data subsets.

3.2 Population-wide genetic diversity and isotype-1 β-tubulin resistance allele frequencies

Genome-wide pairwise F_{ST} was low between all populations, based on population differentiation across 43,587-48,227 markers (Table 2). Population differentiation was highest in the pairwise comparisons involving population U2 (**Table 2**). PCA illustrated the genetic similarity of the two drug-treated populations (T1 and T2), with some overlap with U1 but not U2, consistent with the F_{ST} results (**Fig. 1B**). Removing variation on chromosome I (which contains the isotype-1 β-tubulin gene) decreases differentiation between U1 and T1 and T2 (**Fig. 1C**). Although U2 was collected in the Punjab province, it was collected in Sargodha, about 200 km from Lahore, where the other populations were collected. This distance potentially explains the slightly increased levels of differentiation between U2 and the other populations (**Fig. 1A**).

Of the isotype-1 β-tubulin SNPs previously associated with BZ resistance in *H. contortus,* only 200Y (TTC > T**A**C) was detected in the populations after genotyping by amplicon sequencing. The frequencies of F200Y in the U1, U2, T1, and T2 populations were 2.5%, 12.5%, 100%, and 92.5%, respectively, consistent with their drug treatment history and indicative of their benzimidazole-resistance status (**Fig. 1A inset**). The most common resistance isotype-1 β-tubulin haplotypes are present at high frequency in both T1 and T2 populations (**Fig. 1D**). The isotype-2 β-tubulin gene was also genotyped by

amplicon sequencing for all individual worms to identify potential candidate resistance mutations at codons 167, 198, and 200. All worms were homozygous for "susceptible alleles" (198E, and 200F), except for one worm in population U1 which was heterozygous for a F200Y (TTC > T**A**C) allele.

Population	**T1**	**T2**	**U1**	**U2**
T1		0.023	0.031	0.047
T2			0.033	0.046
U1				0.037

Table 2. Pairwise F_{ST} values between populations.

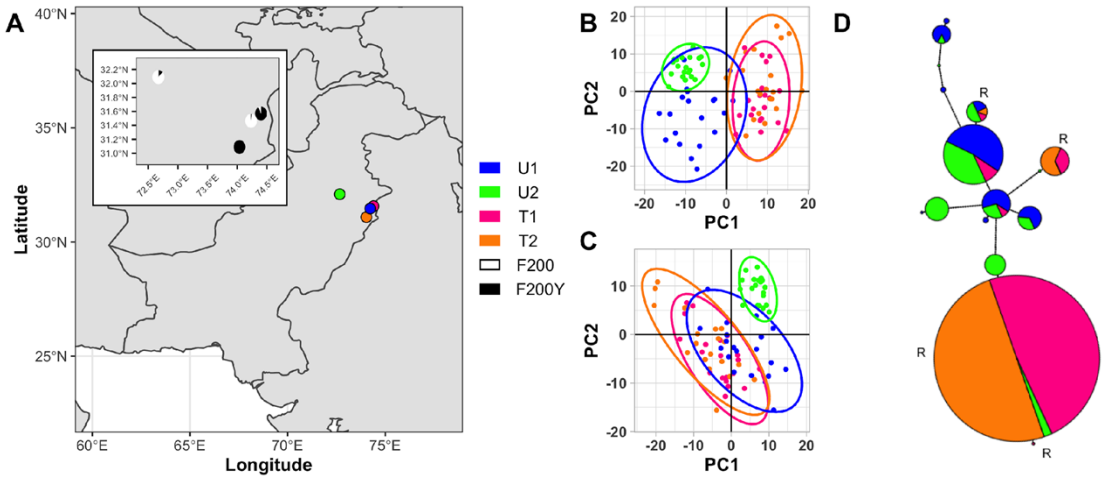

Fig. 1. A) Sampling location, inset shows the percentage of P200Y mutations of the isotype-1 β-tubulin locus. **B)** PCA plot comparing SNP data of the four populations. **C)** PCA plot comparing SNP data of the four populations without chromosome I. **D)** Haplotype network of amplified sequence variants (ASVs) of the isotype-1 beta-tubulin locus amplicon data.

3.3 The major region of genetic differentiation between treated and non-treated populations centers around the isotype-1 β-tubulin locus on chromosome I

All four pairwise genome-wide comparisons between a treated and an untreated population showed a major region of elevated F_{ST} on chromosome I, centered around the isotype-1 β-tubulin locus (chromosome I, around 7.0Mb, **Fig. 2A**). The size of this region depended on the particular pairwise comparison but was between 3.5 and 10 Mb. Pairwise comparisons between either the two treated or two untreated populations did not have an elevated F_{ST} in this region (**Fig. 2A**). In addition, a weaker region of elevated F_{ST} was centred on the isotype-2 β-tubulin locus for the T2-U1 and T2-U2 but not the T1-U1 and T1-U2 pairwise comparisons.

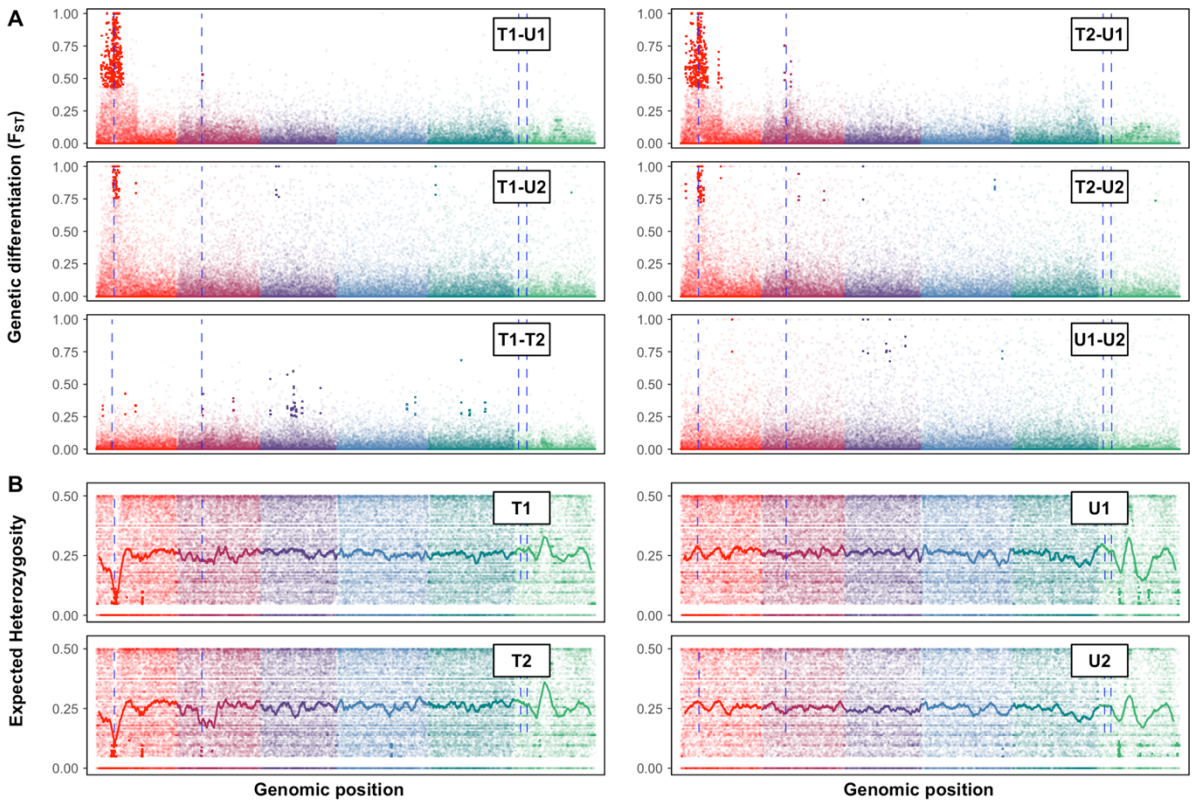

Fig. 2. A) Genome-wide pairwise F_{ST} **B)** Genome-wide expected heterozygosity estimates per population. Chromosomes are indicated by different colors and ordered from I to V with X on the far right in green. Bolded data points represent loci with an F_{ST} significantly different from background noise, as described by

391 the top candidate test [71] and the Materials and Methods. The blue lines indicate the location of isotype-1
392 β-tubulin on chromosome I, isotype-2 β-tubulin on chromosome II, and isotype-3 and 4 β-tubulin on
393 chromosome X. The solid line in B is the rolling average across 500 RAD loci.
394

395 **3.4 Patterns of expected heterozygosity and linkage disequilibrium indicate a**
396 **strong signature of selection around the isotype-1 β-tubulin locus and identify a**
397 **second locus under selection on chromosome I**

398 In the treated but not the untreated populations, the region surrounding the
399 isotype-1 β-tubulin gene, which is located at 7,027,095 Mb on chromosome I, has
400 distinctly reduced expected heterozygosity (position range 5,708,200 Mb - 8,113,930
401 Mb) and elevated linkage disequilibrium (position range 6,075,000 Mb – 8,325,000 Mb)
402 (**Fig. 2B and Fig. 3**). There is a second region on chromosome 1 with a distinct decrease
403 in expected heterozygosity (position range 26,568,594 Mb - 26,676,298 Mb) and an
404 elevated linkage disequilibrium (position range 26,025,000 Mb - 26,625,000 Mb) (**Fig. 3**).
405 A third region on chromosome 2 has a region of reduced expected heterozygosity in
406 treated population T2 (position range 13,221,871 Mb – 13,270,135 Mb) close to the
407 isotype-2 β-tubulin gene (which is located at 13,433,296 Mb) (**Fig. 2B**). Linkage
408 disequilibrium around the third region is not elevated (**Fig. S2**).

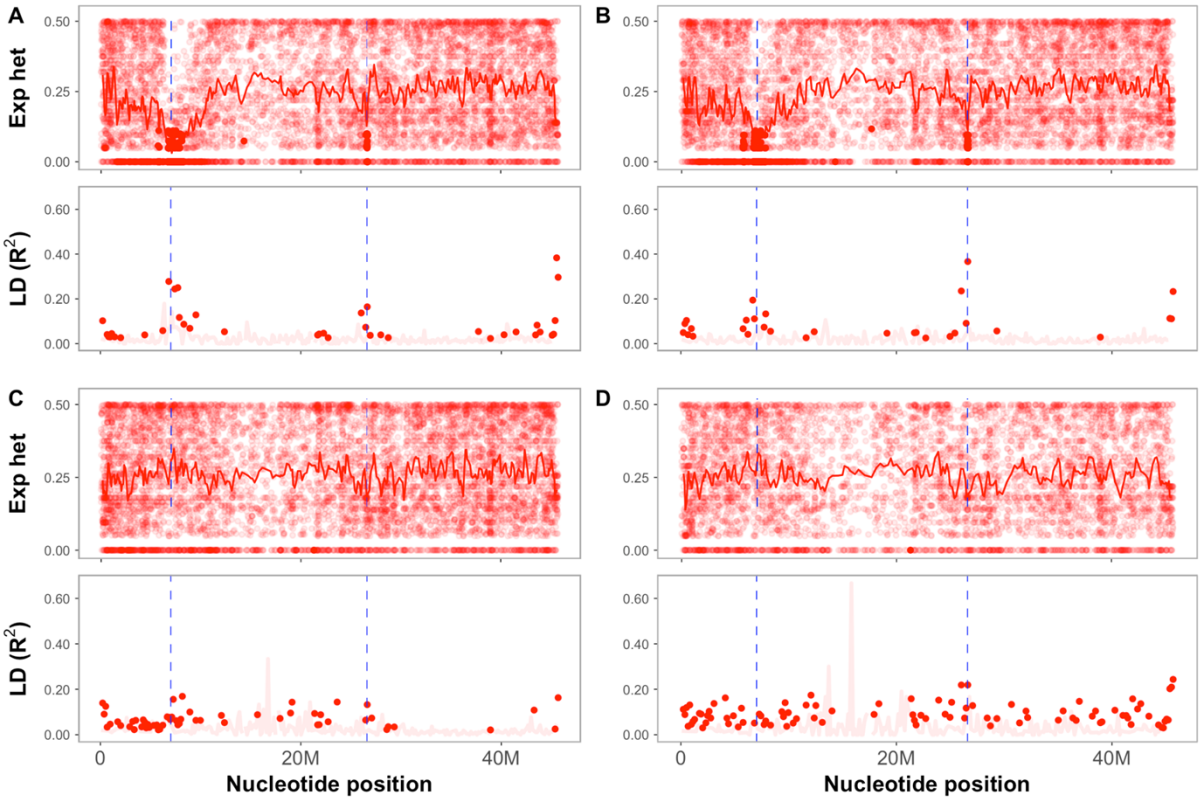

Fig. 3. Chromosome I expected heterozygosity (Exp het), and linkage disequilibrium (R^2 **A)** Population T1 **B)** Population T2 **C)** Population U1, and **D)** Population U2. Bolded data points represent loci significantly different from background noise, as described by the top candidate test [71] and the Materials and Methods. Opaque points and lines indicate non-significant loci. The two dashed lines indicate isotype-1 β-tubulin (left, around 7027095 Mb) and the second locus on chromosome I (right, around 26596797 Mb). The solid line in expected heterozygosity plots is the rolling average across 500 RAD-loci.

3.5 Effect of sequencing depth, sample size, population structure, and genome assembly quality on the signature of selection detected by F_{ST}

We investigated the amount of RADseq data required to detect the signature of selection, as measured by the elevation in F_{ST} by randomly sub-sampling the dataset. As expected, decreasing the numbers of individuals per population in the analysis decreases the signal of elevated F_{ST} around the isotype-1 β-tubulin gene. However, the signal remains detectable even when data from as few as five individual worms are included in the analysis, especially for the comparisons between a treated population (T1 or T2) and

the U1 untreated population (**Fig. 4A**, **Fig. S3**). Decreasing the number of reads per individual only minimally affected the elevation of F_{ST} around the isotype-1 β-tubulin locus until the coverage drops below 3.3X (read depth 0.5M), particularly in comparisons with the more genetically divergent U2 population (**Table 1, Fig. 4B**, **Fig. S4**).

3.6 Gene models within the regions defined by the signals of selection in populations T1 and T2.

With the chromosomal-scale reference genome, the RADseq data was used to detect genomic regions with signatures of selection shared by the treated populations. With the gene annotations that accompany this genome, we further explored these regions for genes related to BZ resistance. We selected three regions of interest based on the results of the four estimates. The narrowest ranges for these three regions, based on an intersection of the detected ranges in all comparisons are (1) chromosome I: 6,450,000-8,113,930 Mb; (2) chromosome I: 26,568,594-26,625,000 Mb; and (3) chromosome II: 13,221,871-13,270,135 Mb. In total, these regions contain 161, five, and one gene(s), respectively (**S2 File**). As discussed, region 1 contains isotype-1 β-tubulin (HCON_00005260). Region 2 contains five genes; HCON_00018620, HCON_00018630, HCON_00018640, HCON_00018650, and HCON_00018660. Region 3, which is based on diversity estimates only, contains a single gene (HCON_00043570) which is predicted to be a hyaluronidase.

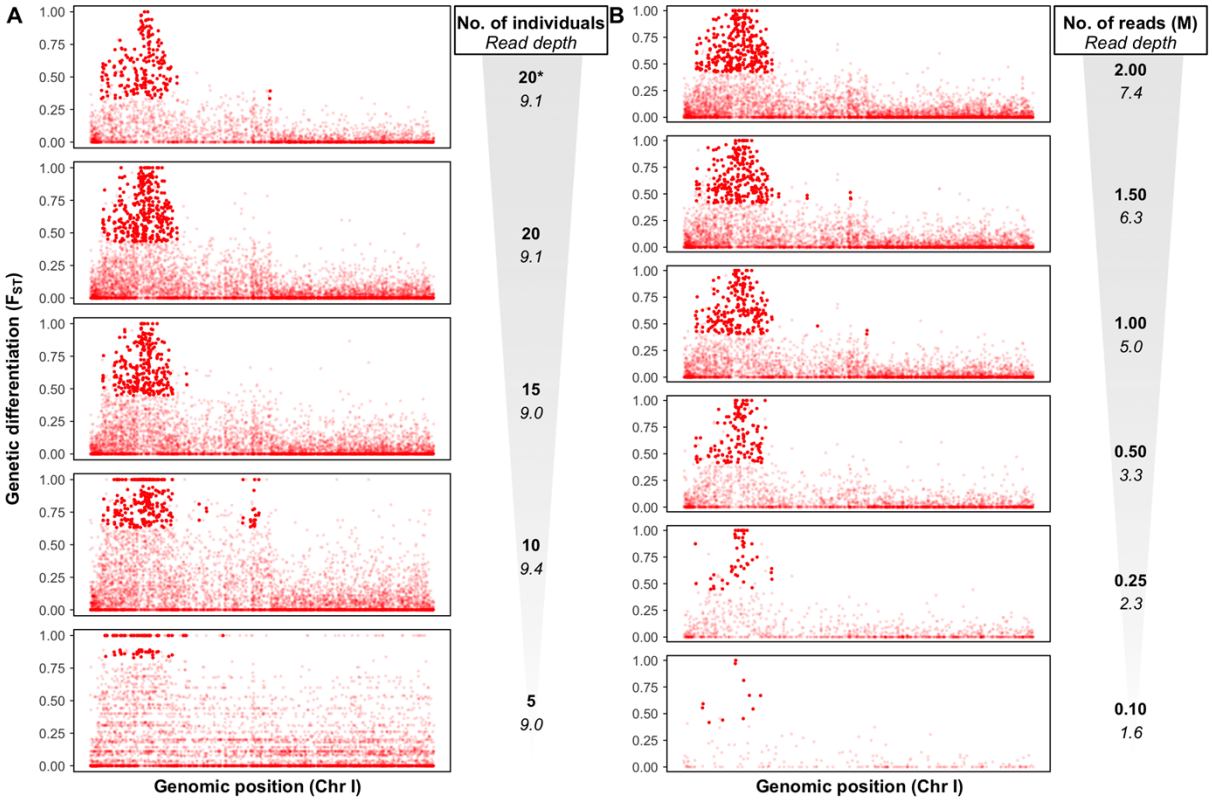

Fig. 4. F_{ST} analysis between T1 and U1. Each dot represents a variable locus between the paired populations, solid colors represent loci with an F_{ST} significantly above background noise, as described by the top candidate method [71] **A)** F_{ST} estimates across chromosome I with decreasing numbers of individuals per population. 20*; stringent, variants are called when represented in 75% of individuals per population. 20, 15, 10 and 5; variant data present for 50% of the individuals per population, 20-15-10-5 **B)** F_{ST} estimates across chromosome I for decreasing numbers of reads per individual sample. Variant data had to be present in 50% of the individuals per population, but only the indicated number of reads was used per sample. Read depth indicated under read count.

3.7 Ability to detect signatures of selection using pooled sequence analysis

We compared the ability of a pool-seq approach to detect the signature of selection to that of the single worm approach by pooling the individual worm RADseq data for each population. Using Popoolation2, 32,234 SNPs were identified as differential in at least one pairwise population comparison (**Fig. 5C**, **Fig. S5,** [74]). These pairwise comparisons show a peak in F_{ST} around the isotype-1 β-tubulin gene, especially for comparisons

between both treated populations and U1, with a less clear peak in the comparison with U2. Diversity (π) was determined with all available SNP data, with 27,719-113,917 estimates across the genome. These loci were then analysed with Popoolation using the loci as a gene map, with 5,764-14,156 π estimates per population. None of the regions under selection as detected with individual samples were identified with diversity estimates in the pooled analysis (**Fig. S6**).

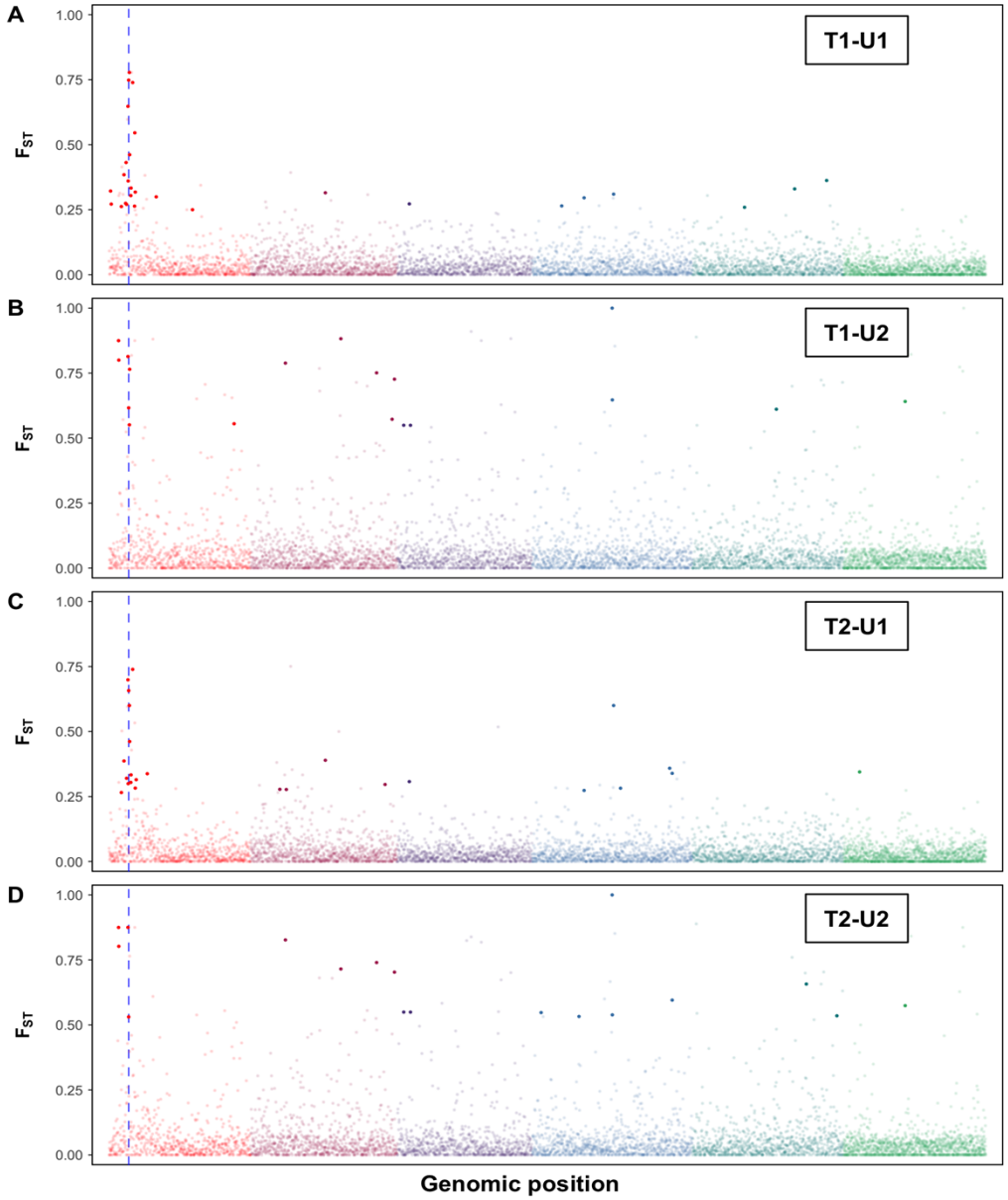

Fig. 5. Pairwise F_{ST} analysis between pooled treated and untreated populations. Each dot represents a variable locus between the paired populations, solid colors represent loci with a F_{ST} estimate, as described by the top candidate method [71]. The dashed blue line indicates the position of isotype-1 beta-tubulin.

4. DISCUSSION

The small ruminant gastrointestinal parasite *H. contortus* is the leading parasitic nematode model used for anthelmintic resistance research [9,76]. The primary aim of this study was to investigate the number, genomic location(s), and characteristics of the signatures of selection specifically associated with long-term routine use of benzimidazole drugs in *H. contortus* in the field. The widespread use of anthelmintic drugs in livestock, together with high parasite migration as a result of animal movement, makes it difficult to find *H. contortus* populations that have either been subject to selection with a single drug class or have not been exposed to drug selection at all. However, our previous work identified two *H. contortus* populations from government farms in Pakistan (designated T1 and T2 in this paper) that have been subjected to intensive selection pressure by benzimidazoles over many years, but not to selection by other drug classes [52]. In the present work we have compared these with two *H. contortus* populations from rural sheep flocks from the same region (designated U1 and U2 in this paper) that are likely to have had minimal exposure to drug selection [52]. Amplicon sequencing confirmed a high frequency of the previously known F200Y isotype-1 β-tubulin benzimidazole resistance mutation in the two selected populations (T1 and T2) and a very low frequency in the two unselected populations (U1 and U2), consistent with the respective drug selection histories of the populations (**Fig. 1A**). Genome-wide pairwise F_{ST} analysis of 43,587-48,227 RADseq markers confirmed low levels of genetic differentiation between the four populations with U2 being the most differentiated. Consequently, these populations were particularly suitable to investigate the number, genomic location(s) and characteristics of the signatures of selection specifically associated with long-term routine use of benzimidazole drugs in the field. In this paper, we have presented a reduced representation genome-wide approach, in which

a large panel of RADseq markers were mapped to the recently completed *H. contortus* chromosomal-scale genome assembly, to identify the main genetic loci under selection by long term use of benzimidazoles these *H. contortus* field populations (Doyle et al., 2020). A single worm genotyping approach was chosen, together with study populations from the same geographical region and minimal genetic differentiation, to maximize sensitivity for detecting signatures of selection in parasite field populations. This experimental design also allowed us to sub-sample in different ways and pool our data in order to investigate how the sensitivity of signal detection varied in terms of sample number and read depth.

4.1 The major genomic signature of selection associated with benzimidazole selection in two *H. contortus* field populations surrounds the isotype-1 β-tubulin locus

There is a substantive body of work indicating the importance of a number of non-synonymous mutations in the isotype-1 β-tubulin gene in benzimidazole resistance in *H. contortus* and related ruminant gastrointestinal nematodes of the superfamily trichostrongyloidea [reviewed in 8,77]. This work includes studies that show evidence of selection and high frequencies of codon F167Y, E198A, E198L, and F200Y mutations in benzimidazole resistant parasite populations, and a recent CRISPR/CAS9 reverse genetic study showed that these specific substitutions in the *C. elegans ben-1* β-tubulin gene are sufficient to confer benzimidazole resistance without a fitness cost [78]. There is also evidence to suggest that, although these isotype-1 β-tubulin gene mutations are important determinants of benzimidazole resistance, other loci may also play a role. For example, in *H. contortus,* deletion of isotype-2 β-tubulin gene [31] and increased levels of benzimidazole glycosidation due to increased UDP-glucuronosyltransferase (UGT)

521 expression [79] have been suggested to contribute to resistance. In addition, in *C. elegans,*
522 a QTL on chromosome IV that does not map to the *ben-1* locus, has been shown to underlie
523 the benzimidazole resistance phenotype of some field populations [80], as well as a QTL
524 on the X chromosome [37]. Whole genome-sequencing of adult *H. contortus* worms from
525 different regions of the world confirms historical selection around the isotype-1 β-tubulin
526 locus in addition to many other loci with signatures of selection [38]. Consequently,
527 although isotype-1 β-tubulin gene mutations are clearly important causes of
528 benzimidazole resistance in *H. contortus,* their relative importance with respect to other
529 genetic loci remains a major knowledge gap.

530 The genome-wide data presented here revealed that the dominant signal of
531 selection for the two different benzimidazole selected *H. contortus* field populations (T1
532 and T2) was centred on the isotype-1 β-tubulin locus on the left arm of chromosome I. No
533 such signal was present in the two populations with a history of little or no benzimidazole
534 selection (U1 and U2). The selection signal is clearly defined by a region of elevated F_{ST} in
535 pairwise comparisons of the two treated with the two non-treated populations and by
536 reduced expected heterozygosity and elevated linkage disequilibrium in the selected but
537 not the unselected populations. The region of elevated F_{ST} extends across a broad region
538 of the left arm of chromosome I (1,663,419 – 11,549,932) but the regions of reduced
539 nucleotide diversity and elevated linkage disequilibrium are much narrower (π:
540 5,708,200 – 8,113,930; LD: 6,450,000 – 8,325,000). The isotype-1 β-tubulin gene is
541 located centrally in these regions around position 7,027,095. Overall, these results
542 strongly suggest that the isotype-1 β-tubulin gene is the single most important
543 benzimidazole resistance locus in the *H. contortus* genome of field populations on both
544 government farms examined in this study.

The T1 and T2 *H. contortus* populations may not be completely independent since the flocks were founded by animals from the same region 30 years ago and there may have been some historical animal movement between the farms during the intervening period [52]. Interestingly, the same major isotype-1 β-tubulin resistance haplotypes were detected by short-read amplicon sequencing (**Fig. 1D**). Also, pairwise F_{ST} values of the RADseq markers surrounding the isotype-1 β-tubulin locus were not elevated in pairwise comparisons between the T1 and T2 populations. Together these suggest the same resistance alleles are present in the two populations and so could have common origins. Nevertheless, these two populations have been independently selected by benzimidazole treatment for around 30 years and our data reveals the region surrounding the isotype-1 β-tubulin locus is the most strongly selected in both cases.

4.2 Additional genomic regions show signatures of selection after benzimidazole treatments

In addition to the predominant selection signature surrounding the isotype-1 β-tubulin locus in both the T1 and T2 populations, the ddRADseq genome-wide scans detected at least one other region of interest towards the middle of chromosome I (26,568,594 – 26,625,000). This region had a clear signature of selection – as defined by increased population differentiation, reduction in nucleotide diversity and elevated linkage disequilibrium – in the treated but not the untreated populations (**Fig. 2**, **Fig. 3A,B**). This signal is much narrower than that surrounding the isotype-1 β-tubulin locus but is clearly defined by three different measures of selection in both treated populations and not in the two untreated populations. The region encompasses only five genes and may contain a novel drug resistance locus, albeit likely of secondary importance to the isotype-1 β-tubulin gene. Gene ontology revealed no specific functions for the five genes

570 identified in the most narrow region. But a gene (HCON_00018690) with an ABC
571 transporter domain is located close to this region (26,693,160-26,714,246 Mb). Deletion
572 or increased expression of ABC transporters have previously been associated with
573 anthelmintic drug susceptibility or resistance in nematodes, respectively [81,82].

574 A second region of potential interest is a small region encompassing one gene on
575 chromosome II (**Fig. 2**; region: 13,221,871 – 13,270,135). This region is one bin removed
576 from the bin containing isotype-2 β-tubulin (13,433,296 Mb) and was marked by elevated
577 F_{ST} between both the treated T1 and T2 isolates and U1, but not U2 (**Fig. 2A**). Additionally,
578 this region has reduced nucleotide diversity in T2, but not the other populations (**Fig.**
579 **2B**). Short-read amplicon sequencing of a region encompassing codons 167, 198 and 200
580 of the isotype-2 β-tubulin gene in the T1 and T2 populations did not detect F167Y, E198A,
581 E198L or F200Y candidate resistance mutation in this gene. As a result, the potential role
582 of variation in the isotype-2 β-tubulin gene with regards to resistance in these
583 populations remains unclear.

584

585 **4.3 ddRADseq marker panel development for genome-wide analysis**

586 In this work we explored the utility of a reduced representation genotyping
587 approach, ddRADseq, for detecting signatures of drug selection in a parasitic nematode
588 genome because of its potential cost effectiveness in large scale field studies, even for
589 laboratories with relatively limited budgets. The challenges and opportunities of RADseq
590 to detect loci associated with adaptation have recently been discussed extensively [83–
591 86]. The extent of linkage disequilibrium (LD), the availability of a reference genome, and
592 the resulting marker density needed for detection of adaptation are the main challenges
593 to address. For this study, we generated ddRADseq data from 20 individual worms from
594 each of four populations, following whole genome amplification.

595 To date, most published applications of ddRADseq on metazoan organisms have
596 involved relatively small marker panels (up to 1 marker per ~20,900 kb according to a
597 recent review of SNP recovery in aquaculture species [87]) and have successfully been
598 used to study population structure. With regards to genome-wide scans to look for
599 signatures of selection, a major challenge is to develop sufficiently dense marker panels
600 across the genome. For most non-model organisms the extent of LD is unknown,
601 complicating the prediction of the number of markers needed to detect signatures of
602 selection. Another challenge is allelic dropout, which occurs as a result of restriction site
603 heterogeneity, and limits the number of recovered loci that are shared between
604 populations [88,89]. For *H. contortus*, allelic dropout is likely as a result of of its extremely
605 high levels of genetic diversity (reviewed in Gilleard and Redman [90]). We performed an
606 *in silico* digestion with the previously published draft reference genome ([62],
607 GCA_000469685.1) to choose a combination of restriction enzymes and fragment sizes to
608 produce a panel of ~50k markers that we could recover from at least three of the four
609 populations. At an average of 1 marker per 3kb this represents one of the densest
610 ddRADseq marker sets produced for a metazoan organism to date. Two recent papers
611 have described RADseq SNP marker panels for *H. contortus* with 2,667 [91] and 82,271
612 [92] markers, respectively. Although the latter study reported a large number of SNPs,
613 this is not comparable to the present study because they reported multiple SNPs per
614 RADseq marker and genotyped pooled genomic DNA to study genetic diversity within
615 and between five populations from different countries.

4.4. Considerations to optimize genome-wide approaches to identify signatures of selection in parasitic nematode field populations

Our study and methodology was designed to maximize the sensitivity of detection of the genomic signals of selection associated with benzimidazole treatments of *H. contortus* in the field. The single worm ddRADseq dataset we generated comprised an average sequence depth of >9 reads per marker per sample which we examined in a number of ways to investigate how the sensitivity of detection, based on pairwise F_{ST} analysis and expected heterozygosity, was impacted by the number of individuals, total read depth and reference genome quality. We also investigated the ability of pooled sequence data to detect the signals.

H. contortus has substantial population structure between countries and a low but discernible population structure within countries [reviewed by 93]. The identification of genomic signatures of selection in specific regions of the genome can be confounded by population structure [94]. Consequently, the four study populations were taken from the same geographical region to limit the extent of between-population genetic differentiation across the genome. The expected limited differentiation was confirmed by the genome-wide F_{ST} values of <0.05 for all pairwise comparisons (**Table 2**). Further, most of genetic differentiation between the populations was accounted for by chromosome I suggesting it was mainly the result of selection at the isotype-1 β-tubulin locus (**Fig. 1B,C**). Population U2 appeared to be the most genetically distinct of the four populations if the variation on chromosome I is ignored (**Fig. 1C**). There are two alternative explanations for this: (i) genuine genetic differentiation of this population with respect to the others, especially given the relative distance of the sampling location of U2 in Sargodha compared with those of T1, T2 and U1 in Lahore (all four located in the Lahore region), or (ii) a technical artifact as a result of generating the RADseq data for

641 this population on a separate sequencing run as a result of initial poor read recovery of
642 that sample. Fragment size selection during library preparation could have affected the
643 recovered markers. However, based on the markers shared across all populations, this is
644 not a significant concern. Whatever the cause, the strength of the signature of selection
645 around the isotype-1 β-tubulin locus was much lower in pairwise F_{ST} comparisons using
646 the U2 than the U1 population (**Fig. 2**). This illustrates the importance of careful matching
647 of parasite populations being compared and controlling of experimental procedures to
648 minimize genome-wide differentiation and so maximize the sensitivity of detection.
649 The signature of selection around isotype-1 β-tubulin could be detected with
650 RADseq data from as few as 5 individual worms per population and as little as two reads
651 per marker when 20 individuals were used (**Table 1 and Fig. 4**). Based on the data
652 presented here, aiming for a 5-6.5X coverage of 1.1 to 1.4% of the genome after data
653 cleaning should allow detection of strong signatures of selection. The number of markers
654 expected based on the species of interest and enzyme combination can be calculated *in*
655 *silico* (Lepais and Weir 2014; Mora-Márquez et al. 2017). For expected heterozygosity,
656 the signatures of selection disappear earlier than for F_{ST}; at less than 15 individuals and
657 between 1 and 0.5M reads (**Figs. S7 and S8**). An advantage of expected heterozygosity is
658 that populations do not have to be as carefully matched.
659 Draft reference genome sequences are becoming available for an increasing
660 number of parasitic nematode genomes (50 Helminth Genomes Project, 50HGP;
661 http://www.sanger.ac.uk/science/collaboration/50hgp). However, there are still very
662 few chromosomal-scale reference genome assemblies for these organisms [but see 66].
663 Consequently, genome-wide analyses in parasitic nematodes are generally undertaken
664 using draft reference genomes with limited contiguity [95–97]. In order to simulate this
665 situation, we mapped the full set of RADseq read data against a previously published *H.*

contortus draft reference genome ([62], GCA_000469685.1). Numerous scaffolds with elevated F_{ST} relative to the rest of the genome were identified as showing a signal of selection (**Fig. S9**). In the absence of positional information for the different genomic scaffolds, this data could be potentially misinterpreted as evidence of multiple signatures of selection being present in the genome when in fact it is due to the hitchhiking effect of a single large selective sweep around the isotype-1 β-tubulin locus. This is an important point to consider when working with draft reference genomes that still have relatively poor contiguity.

We also compared the detection of genomic selection signatures using pooled versus single worm RADseq data. We pooled data from individual samples by merging reads from all samples per population into one sample and selecting enough reads for 100 x coverage. F_{ST} is high around the isotype-1 β-tubulin gene in comparisons with U1, but it is not significant in comparisons with U2 (**Fig. S5**). This result indicates the limited power of F_{ST} estimates in pooled populations if there is some population structure present. For expected heterozygosity no evidence of selective sweeps was detected using the pooled data (**Fig. S6**). The use of pooled data for genome-wide scans can be improved by increasing sample counts [98] but the ability to do this will depend on the number of well-defined isolates and the budget available. Our data provides a framework to balance the relative merits of sequencing a smaller number of worms at low depth from a small number of populations compared with undertaking Poolseq using greater sequencing depth on a greater number of populations.

In summary, this study illustrates the power of genome-wide approaches to identify signatures of selection associated with long term anthelmintic treatment of parasitic nematodes in the field. It has shown the critical importance of minimizing the population structure when selected and unselected populations are being compared and

691 the greater sensitivity achieved by single worm sequencing compared to Poolseq
692 approaches. We have shown that the isotype-1 β-tubulin gene is quantitatively by far the
693 most important benzimidazole resistance locus in the *H. contortus* genome in the two
694 field populations subjected to long term drug treatment. Interestingly however, we have
695 identified two additional loci that likely harbor mutations associated with benzimidazole
696 resistance but are quantitatively of lesser importance than isotype-1 β -tubulin. Further
697 investigation of these regions may reveal new genes associated with benzimidazole
698 resistance.
699

700 **ACKNOWLEDGEMENTS**

701 We would like to thank Stephen Doyle, Axel Martinelli, and the members of the
702 HPI training program for their support during various aspects of the project. We would
703 also like to thank WormBase ParasSite for hosting the *H. contortus* genome and
704 annotation. For the purpose of Open Access, the author has applied a CC BY public
705 copyright license to any Author Accepted Manuscript version arising from this
706 submission.
707

SUPPLEMENTARY MATERIAL

S1 File Sequencing primers

S2 File Genes in regions of interest

S1 Fig. Marker density across the genome

S2 Fig. LD for all populations

S3 Fig. Population differentiation with reducing number of individuals; all populations and the whole genome

S4 Fig. Population differentiation with reducing read count; all populations and the whole genome

S5 Fig. Population differentiation with pooled data

S6 Fig. Diversity estimates with pooled data

S7 Fig. Diversity estimates with reducing read count

S8 Fig. Diversity estimates with reducing number of individuals

S9 Fig. All pairwise comparisons with the draft genome

REFERENCES

1. Rose Vineer H, Morgan ER, Hertzberg H, Bartley DJ, Bosco A, Charlier J, et al. Increasing importance of anthelmintic resistance in European livestock: creation and meta-analysis of an open database. Parasite. 2020;27: 69.

2. Mutombo PN, Man NWY, Nejsum P, Ricketson R, Gordon CA, Robertson G, et al. Diagnosis and drug resistance of human soil-transmitted helminth infections: A public health perspective. Adv Parasitol. 2019;104: 247–326.

3. Kaplan RM, Vidyashankar AN. An inconvenient truth: global worming and anthelmintic resistance. Vet Parasitol. 2012;186: 70–78.

4. Furtado LFV, Medeiros C da S, Zuccherato LW, Alves WP, de Oliveira VNGM, da Silva VJ, et al. First identification of the benzimidazole resistance-associated F200Y SNP in the beta-tubulin gene in Ascaris lumbricoides. PLoS One. 2019;14: e0224108.

5. Orr AR, Quagraine JE, Suwondo P, George S, Harrison LM, Dornas FP, et al. Genetic Markers of Benzimidazole Resistance among Human Hookworms (Necator americanus) in Kintampo North Municipality, Ghana. Am J Trop Med Hyg. 2019;100: 351–356.

6. Krücken J, Fraundorfer K, Mugisha JC, Ramünke S, Sifft KC, Geus D, et al. Reduced efficacy of albendazole against Ascaris lumbricoides in Rwandan schoolchildren. Int J Parasitol Drugs Drug Resist. 2017;7: 262–271.

7. Tinkler SH. Preventive chemotherapy and anthelmintic resistance of soil-transmitted helminths - Can we learn nothing from veterinary medicine? One Health. 2020;9: 100106.

8. Kotze AC, Gilleard JS, Doyle SR, Prichard RK. Challenges and opportunities for the adoption of molecular diagnostics for anthelmintic resistance. Int J Parasitol Drugs Drug Resist. 2020;14: 264–273.

9. Gilleard JS. Haemonchus contortus as a paradigm and model to study anthelmintic drug resistance. Parasitology. 2013;140: 1506–1522.

10. Kwa MSG, Veenstra JG, Roos MH. Benzimidazole resistance in Haemonchus contortus is correlated with a conserved mutation at amino acid 200 in β-tubulin isotype 1. Mol Biochem Parasitol. 1994;63: 299–303.

11. Rufener L, Mäser P, Roditi I, Kaminsky R. Haemonchus contortus acetylcholine receptors of the DEG-3 subfamily and their role in sensitivity to monepantel. PLoS Pathog. 2009;5: e1000380.

12. Urdaneta-Marquez L, Bae SH, Janukavicius P, Beech R, Dent J, Prichard R. 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. 2014;44: 1063–1071.

13. McCavera S, Rogers AT, Yates DM, Woods DJ, Wolstenholme AJ. An ivermectin-

sensitive glutamate-gated chloride channel from the parasitic nematode Haemonchus contortus. Mol Pharmacol. 2009;75: 1347–1355.

14. Le Jambre LF, Gill JH, Lenane IJ, Baker P. Inheritance of avermectin resistance in Haemonchus contortus. Int J Parasitol. 2000;30: 105–111.

15. Kaminsky R, Ducray P, Jung M, Clover R, Rufener L, Bouvier J, et al. A new class of anthelmintics effective against drug-resistant nematodes. Nature. 2008;452: 176–180.

16. Guest M, Bull K, Walker RJ, Amliwala K, O'Connor V, Harder A, et al. The calcium-activated potassium channel, SLO-1, is required for the action of the novel cyclo-octadepsipeptide anthelmintic, emodepside, in Caenorhabditis elegans. Int J Parasitol. 2007;37: 1577–1588.

17. Wit J, Dilks CM, Andersen EC. Complementary Approaches with Free-living and Parasitic Nematodes to Understanding Anthelmintic Resistance. Trends Parasitol. 2021;37: 240–250.

18. Rezansoff AM, Laing R, Martinelli A, Stasiuk S, Redman E, Bartley D, et al. The confounding effects of high genetic diversity on the determination and interpretation of differential gene expression analysis in the parasitic nematode Haemonchus contortus. Int J Parasitol. 2019;49: 847–858.

19. Prichard RK, Roulet A. ABC transporters and beta-tubulin in macrocyclic lactone resistance: prospects for marker development. Parasitology. 2007;134: 1123–1132.

20. Glendinning SK, Buckingham SD, Sattelle DB, Wonnacott S, Wolstenholme AJ. Glutamate-gated chloride channels of Haemonchus contortus restore drug sensitivity to ivermectin resistant Caenorhabditis elegans. PLoS One. 2011;6: e22390.

21. Janssen IJI, Krücken J, Demeler J, Basiaga M, Kornaś S, von Samson-Himmelstjerna G. Genetic variants and increased expression of Parascaris equorum P-glycoprotein-11 in populations with decreased ivermectin susceptibility. PLoS One. 2013;8: e61635.

22. Doyle SR, Cotton JA. Genome-wide approaches to investigate anthelmintic resistance. Trends Parasitol. 2019;35: 289–301.

23. Valentim CLL, Cioli D, Chevalier FD, Cao X, Taylor AB, Holloway SP, et al. Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. Science. 2013;342: 1385–1389.

24. Anderson TJC, LoVerde PT, Le Clec'h W, Chevalier FD. Genetic Crosses and Linkage Mapping in Schistosome Parasites. Trends Parasitol. 2018;34: 982–996.

25. Redman E, Grillo V, Saunders G, Packard E, Jackson F, Berriman M, et al. Genetics of mating and sex determination in the parasitic nematode Haemonchus contortus. Genetics. 2008;180: 1877–1887.

26. Redman E, Sargison N, Whitelaw F, Jackson F, Morrison A, Bartley DJ, et al. Introgression of ivermectin resistance genes into a susceptible Haemonchus contortus strain by multiple backcrossing. PLoS Pathog. 2012;8: e1002534.

27. Hunt PW, Kotze AC, Knox MR, Anderson LJ, McNally J, LE Jambre LF. The use of DNA markers to map anthelmintic resistance loci in an intraspecific cross of Haemonchus contortus. Parasitology. 2010;137: 705–717.

28. Doyle SR, Illingworth CJR, Laing R, Bartley DJ, Redman E, Martinelli A, et al. Population genomic and evolutionary modelling analyses reveal a single major QTL for ivermectin drug resistance in the pathogenic nematode, Haemonchus contortus. BMC Genomics. 2019;20: 218.

29. Doyle SR, Laing R, Bartley D, Morrison A, Holroyd N, Maitland K, et al. Genomic landscape of drug response reveals novel mediators of anthelmintic resistance. bioRxiv. 2022. doi:10.1101/2021.11.12.465712

30. Laing R, Doyle SR, McIntyre J, Maitland K, Morrison A, Bartley DJ, et al. Transcriptomic analyses implicate neuronal plasticity and chloride homeostasis in ivermectin resistance and recovery in a parasitic nematode. bioRxiv. 2021. p. 2021.11.12.468372. doi:10.1101/2021.11.12.468372

31. Kwa MSG, Veenstra JG, Roos MH. Molecular characterisation of β-tubulin genes present in benzimidazole-resistant populations of Haemonchus contortus. Mol Biochem Parasitol. 1993;60: 133–143.

32. Kwa MSG, Veenstra JG, Van Dijk M, Roos MH. β-Tubulin Genes from the Parasitic NematodeHaemonchus contortusModulate Drug Resistance inCaenorhabditis elegans. J Mol Biol. 1995;246: 500–510.

33. Dilks CM, Hahnel SR, Sheng Q, Long L, McGrath PT, Andersen EC. Quantitative benzimidazole resistance and fitness effects of parasitic nematode beta-tubulin alleles. Int J Parasitol Drugs Drug Resist. 2020;14: 28–36.

34. Beech RN, Prichard RK, Scott ME. Genetic variability of the beta-tubulin genes in benzimidazole-susceptible and -resistant strains of Haemonchus contortus. Genetics. 1994;138: 103–110.

35. Prichard R. Genetic variability following selection of Haemonchus contortus with anthelmintics. Trends Parasitol. 2001;17: 445–453.

36. Dilks CM, Koury EJ, Buchanan CM, Andersen EC. Newly identified parasitic nematode beta-tubulin alleles confer resistance to benzimidazoles. bioRxiv. 2021. p. 2021.07.26.453836. doi:10.1101/2021.07.26.453836

37. Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles. PLoS Pathog. 2018;14: e1007226.

38. Sallé G, Doyle SR, Cortet J, Cabaret J, Berriman M, Holroyd N, et al. The global diversity of Haemonchus contortus is shaped by human intervention and climate.

838 Nat Commun. 2019;10: 4811.

839 39. Wit J, Gilleard JS. Resequencing Helminth Genomes for Population and Genetic
840 Studies. Trends Parasitol. 2017;33: 388–399.

841 40. Díaz-Arce N, Rodríguez-Ezpeleta N. Selecting RAD-Seq Data Analysis Parameters
842 for Population Genetics: The More the Better? Front Genet. 2019;10: 533.

843 41. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power
844 of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17: 81–
845 92.

846 42. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an
847 inexpensive method for de novo SNP discovery and genotyping in model and non-
848 model species. PLoS One. 2012;7: e37135.

849 43. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP
850 discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3:
851 e3376.

852 44. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. Brief Funct
853 Genomics. 2010;9: 416–423.

854 45. Kumar G, Kocour M. Applications of next-generation sequencing in fisheries
855 research: A review. Fish Res. 2017;186: 11–22.

856 46. Settepani V, Schou MF, Greve M, Grinsted L, Bechsgaard J, Bilde T. Evolution of
857 sociality in spiders leads to depleted genomic diversity at both population and
858 species levels. Mol Ecol. 2017;26: 4197–4210.

859 47. Graham CF, Boreham DR, Manzon RG, Stott W, Wilson JY, Somers CM. How "simple"
860 methodological decisions affect interpretation of population structure based on
861 reduced representation library DNA sequencing: A case study using the lake
862 whitefish. PLoS One. 2020;15: e0226608.

863 48. Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, et al.
864 Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic
865 diversity. Nat Genet. 2012;44: 285–290.

866 49. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population
867 genomics of parallel adaptation in threespine stickleback using sequenced RAD
868 tags. PLoS Genet. 2010;6: e1000862.

869 50. Jacobsen MW, Pujolar JM, Bernatchez L, Munch K, Jian J, Niu Y, et al. Genomic
870 footprints of speciation in Atlantic eels (Anguilla anguilla and A. rostrata). Mol Ecol.
871 2014;23: 4785–4798.

872 51. Wright BR, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. A
873 demonstration of conservation genomics for threatened species management. Mol
874 Ecol Resour. 2020;20: 1526–1541.

52. Chaudhry U, Redman EM, Ashraf K, Shabbir MZ, Rashid MI, Ashraf S, et al. Microsatellite marker analysis of Haemonchus contortus populations from Pakistan suggests that frequent benzimidazole drug treatment does not result in a reduction of overall genetic diversity. Parasit Vectors. 2016;9: 349.

53. Bennett AB, Anderson TJC, Barker GC, Michael E, Bundy DAP. Sequence variation in the Trichuris trichiura β-tubulin locus: implications for the development of benzimidazole resistance. Int J Parasitol. 2002;32: 1519–1528.

54. Avramenko RW, Redman EM, Melville L, Bartley Y, Wit J, Queiroz C, et al. Deep amplicon sequencing as a powerful new tool to screen for sequence polymorphisms associated with anthelmintic resistance in parasitic nematode populations. Int J Parasitol. 2019;49: 13–26.

55. Avramenko RW, Redman EM, Lewis R, Yazwinski TA, Wasmuth JD, Gilleard JS. Exploring the Gastrointestinal "Nemabiome": Deep Amplicon Sequencing to Quantify the Species Composition of Parasitic Nematode Communities. PLoS One. 2015;10: e0143559.

56. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75: 7537–7541.

57. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10–12.

58. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13: 581–583.

59. Team RC, Others. R: A language and environment for statistical computing. Vienna, Austria; 2013. Available: http://finzi.psych.upenn.edu/R/library/dplR/doc/intro-dplR.pdf

60. Paradis E. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 2010;26: 419–420.

61. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20: 289–290.

62. Laing R, Kikuchi T, Martinelli A, Tsai IJ, Beech RN, Redman E, et al. The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. Genome Biol. 2013;14: R88.

63. Lepais O, Weir JT. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour. 2014;14: 1314–1321.

64. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22: 3124–3140.

914 65. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FastQC.
915 2010.

916 66. Doyle SR, Tracey A, Laing R, Holroyd N, Bartley D, Bazant W, et al. Genomic and
917 transcriptomic variation defines the chromosome-scale assembly of Haemonchus
918 contortus, a model gastrointestinal worm. Commun Biol. 2020;3: 656.

919 67. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
920 2012;9: 357–359.

921 68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
922 Alignment/Map format and SAMtools. Bioinformatics. 2009. pp. 2078–2079.
923 doi:10.1093/bioinformatics/btp352

924 69. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
925 data. Bioinformatics. 2011;27: 3070–3071.

926 70. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call
927 format data in R. Mol Ecol Resour. 2017;17: 44–53.

928 71. Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, et al.
929 Convergent local adaptation to climate in distantly related conifers. Science.
930 2016;353: 1431–1433.

931 72. Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. Curr
932 Protoc Bioinformatics. 2014;47: 11–12.

933 73. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
934 features. Bioinformatics. 2010;26: 841–842.

935 74. Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, et al.
936 PoPoolation: a toolbox for population genetic analysis of next generation
937 sequencing data from pooled individuals. PLoS One. 2011;6: e15925.

938 75. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation
939 between populations using sequencing of pooled DNA samples (Pool-Seq).
940 Bioinformatics. 2011;27: 3435–3436.

941 76. Wit J, Dilks CM, Andersen EC. Complementary Approaches with Free-living and
942 Parasitic Nematodes to Understanding Anthelmintic Resistance. Trends Parasitol.
943 2021;37: 240–250.

944 77. Gilleard JS. Understanding anthelmintic resistance: the need for genomics and
945 genetics. Int J Parasitol. 2006;36: 1227–1239.

946 78. Dilks CM, Hahnel SR, Sheng Q, Long L, McGrath PT, Andersen EC. Quantitative
947 benzimidazole resistance and fitness effects of parasitic nematode beta-tubulin
948 alleles. Int J Parasitol Drugs Drug Resist. 2020;14: 28–36.

949 79. Matoušková P, Lecová L, Laing R, Dimunová D, Vogel H, Raisová Stuchlíková L, et al.
950 UDP-glycosyltransferase family in Haemonchus contortus: Phylogenetic analysis,

constitutive expression, sex-differences and resistance-related differences. Int J Parasitol Drugs Drug Resist. 2018;8: 420–429.

80. Zamanian M, Cook DE, Zdraljevic S, Brady SC, Lee D, Lee J, et al. Discovery of genomic intervals that underlie nematode responses to benzimidazoles. PLoS Negl Trop Dis. 2018;12: e0006368.

81. Lespine A, Ménez C, Bourguinat C, Prichard RK. P-glycoproteins and other multidrug resistance transporters in the pharmacology of anthelmintics: Prospects for reversing transport-dependent anthelmintic resistance. Int J Parasitol Drugs Drug Resist. 2012;2: 58–75.

82. Langeland A, Jetter H, O'Halloran DM. The diversity of ABC transporter genes across the Phylum Nematoda. Parasitol Int. 2021;83: 102357.

83. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Responsible RAD: Striving for best practices in population genomic studies of adaptation. Molecular ecology resources. Wiley Online Library; 2017. pp. 366–369.

84. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Resour. 2017;17: 142–152.

85. McKinney GJ, Larson WA, Seeb LW, Seeb JE. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). Molecular ecology resources. Wiley Online Library; 2017. pp. 356–361.

86. Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Molecular ecology resources. Wiley Online Library; 2017. pp. 362–365.

87. Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, et al. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020;21: 389–409.

88. Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. Special features of RAD Sequencing data: implications for genotyping. Mol Ecol. 2013;22: 3151–3164.

89. Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE. Demystifying the RAD fad. Mol Ecol. 2014;23: 5937–5942.

90. Gilleard JS, Redman E. Genetic Diversity and Population Structure of Haemonchus contortus. Haemonchus contortus and Haemonchosis – Past, Present and Future Trends. 2016. pp. 31–68. doi:10.1016/bs.apar.2016.02.009

91. Luo X, Shi X, Yuan C, Ai M, Ge C. Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in Haemonchus …. Parasit Vectors. 2017. Available:

https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-016-1959-6

92. Khan S, Zhao X, Hou Y, Yuan C, Li Y, Luo X, et al. Analysis of genome-wide SNPs based on 2b-RAD sequencing of pooled samples reveals signature of selection in different populations of Haemonchus contortus. J Biosci. 2019;44. doi:10.1007/s12038-019-9917-z

93. Gilleard JS, Redman E. Genetic Diversity and Population Structure of Haemonchus contortus. Haemonchus contortus and Haemonchosis – Past, Present and Future Trends. 2016. pp. 31–68. doi:10.1016/bs.apar.2016.02.009

94. Andolfatto P. Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev. 2001;11: 635–641.

95. Choi Y-J, Bisset SA, Doyle SR, Hallsworth-Pepin K, Martin J, Grant WN, et al. Genomic introgression mapping of field-derived multiple-anthelmintic resistance in Teladorsagia circumcincta. PLoS Genet. 2017;13: e1006857.

96. Łopieńska-Biernat E, Paukszto Ł, Jastrzębski JP, Myszczyński K, Polak I, Stryiński R. Genome-wide analysis of Anisakis simplex sensu lato: the role of carbohydrate metabolism genes in the parasite's development. Int J Parasitol. 2019;49: 933–943.

97. Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, et al. The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat Genet. 2016;48: 299–307.

98. Lynch M, Bost D, Wilson S, Maruki T, Harrison S. Population-genetic inference from pooled-sequencing data. Genome Biol Evol. 2014;6: 1210–1218.