Genomic introgression mapping of field-derived multiple-anthelmintic resistance in Teladorsagia circumcincta

Young-Jun Choi  
*Washington University School of Medicine in St. Louis*

Stewart A. Bisset  
*Hopkirk Research Institute, Palmerston North, New Zealand*

Stephen R. Doyle  
*La Trobe University, Melbourne*

Kymberlie Hallsworth-Pepin  
*Washington University School of Medicine in St. Louis*

John Martin  
*Washington University School of Medicine in St. Louis*

*See next page for additional authors*

**Follow this and additional works at:** [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Recommended Citation**  
[https://digitalcommons.wustl.edu/open_access_pubs/6048](https://digitalcommons.wustl.edu/open_access_pubs/6048)
Genomic introgression mapping of field-derived multiple-anthelmintic resistance in *Teladorsagia circumcincta*

Young-Jun Choi1*, Stewart A. Bisset2*, Stephen R. Doyle3*, Kymberlie Hallsworth-Pepin1, John Martin1, Warwick N. Grant2*, Makedonka Mitreva1,4*

1 McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, Missouri, United States of America, 2 AgResearch, Hopkirk Research Institute, Palmerston North, New Zealand, 3 Department of Animal, Plant and Soil Sciences, La Trobe University, Melbourne, Victoria, Australia, 4 Department of Medicine, Washington University School of Medicine, Saint Louis, Missouri, United States of America

* These authors contributed equally to this work.

☯ Current address: Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

* mmitreva@wustl.edu (MM); w.grant@latrobe.edu.au (WNG)

Abstract

Preventive chemotherapy has long been practiced against nematode parasites of livestock, leading to widespread drug resistance, and is increasingly being adopted for eradication of human parasitic nematodes even though it is similarly likely to lead to drug resistance. Given that the genetic architecture of resistance is poorly understood for any nematode, we have analyzed multidrug resistant *Teladorsagia circumcincta*, a major parasite of sheep, as a model for analysis of resistance selection. We introgressed a field-derived multiresistant genotype into a partially inbred susceptible genetic background (through repeated back-crossing and drug selection) and performed genome-wide scans in the backcross progeny and drug-selected F2 populations to identify the major genes responsible for the multidrug resistance. We identified variation linking candidate resistance genes to each drug class. Putative mechanisms included target site polymorphism, changes in likely regulatory regions and copy number variation in efflux transporters. This work elucidates the genetic architecture of multiple anthelmintic resistance in a parasitic nematode for the first time and establishes a framework for future studies of anthelmintic resistance in nematode parasites of humans.

Author summary

*Teladorsagia circumcincta* is an economically significant nematode (roundworm) pathogen affecting sheep and goats in temperate regions of the world. The widespread use of prophylactic treatment has resulted in rapid selection for anthelmintic (anti-worm drug) resistance in this and other species of livestock parasites. The mechanism of resistance is not well understood because most studies have focused on the role of candidate genes using simplistic models of single gene selection, despite evidence that the evolution of...
resistance is more complex. Here, we report on a comprehensive whole-genome analysis that elucidated resistance-associated genes, which was facilitated by developing a pair of \textit{T. circumcincta} strains sharing a largely common genetic background but differing markedly in their susceptibility to anthelmintic drugs. The results show that multiple genetic factors contribute to anthelmintic resistance in a variety of ways, including possible reduction/modulation in target site sensitivity, reduced target site expression, and increased drug efflux, to name a few. This suggests that drug resistance in these parasites is a multifactorial quantitative trait rather than a simple discrete Mendelian character. With this study, we established a genomics-based experimental paradigm for investigating anthelmintic resistance, at a time when its medical importance is rapidly increasing.

Introduction

Anthelmintic resistance is already a global problem for agriculture and a growing concern in relation to human pathogens \cite{1, 2}. In the absence of effective vaccines, treatment and prophylaxis of helminthiases rely on a limited number of chemotherapeutic agents whose efficacy is increasingly undermined by the selection and spread of resistant parasites. Although fundamental to our ability to conserve sensitivity to existing drugs and to design improved interventions, the molecular and population genetic bases of anthelmintic resistance remain inadequately understood \cite{3, 4}. To date, most studies have focused on the role of individual candidate genes such as drug targets or transporters. However, while such studies have been instrumental in identifying some causal genetic variants associated with drug resistance, frequently they have accounted for only a proportion of the drug resistant phenotypes present in the population, suggesting that the trait probably has a complex multi-genic nature \cite{5–7}. Efforts to comprehensively map functional polymorphisms and to clarify the genotype-phenotype relationships in anthelmintic resistance have been challenging, given the relatively poor genetic and experimental tractability of helminth systems, which has impeded genome-wide studies beyond targeted analysis of particular candidate genes \cite{8}.

The problem of anthelmintic resistance is most severe in the trichostrongylid nematodes of livestock and particularly those infecting small ruminants such as sheep and goats \cite{2}. The troubling propensity of these parasites to develop drug resistance has been attributed to their enormous effective population size and the resulting genetic diversity upon which selection is able to act \cite{9, 10} and although variation is a prerequisite for selection, the extreme genetic heterogeneity in parasite populations often confounds the identification or association of genetic components contributing toward anthelmintic resistance. This difficulty is further hampered by factors such as the degree of parasite population connectivity due to parasite and/or host movement \cite{11}, the influence of population size and life history traits on genetic drift within parasite subpopulations, and the variation in local parasite management strategies, all of which likely influence the ability to detect and correctly interpret genetic differentiation between anthelmintic resistant and susceptible parasites \cite{10}. Our approach towards identifying drug resistance associated genes involved the controlled crossing of a multidrug-resistant parasite strain with a characterized susceptible strain followed by repeated backcrossing and drug selection, which resulted in the introgression of resistance associated alleles into a largely susceptible, partially inbred genetic background. By identifying the alleles derived from the original resistant parent in the resulting backcrossed progeny, it was possible to generate a genetic map of resistance loci within the genome. Similar approaches have been used in mapping drug

molecular work undertaken on them, were carried out at AgResearch’s Wallaceville Animal Research Centre and/or Hopkirk Research Institute in New Zealand, with funding from the New Zealand Foundation for Research, Science and Technology (Grant No. C10 616). The genome sequencing and annotation work was funded by US National Institutes of Health (NIH)—National Human Genome Research Institute. Comparative genome analysis was funded by grants NIH—National Institute of Allergy and Infectious Diseases AI081803 and NIH—National Institute of General Medical Sciences GM097435 to M.M. S.R.D. was supported by a Illumina MiSeq grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.
resistance associated loci in *Haemonchus contortus* [12, 13] and elucidating the genetic basis of drug resistance in some trematode parasite species [14].

In this study, we extended our previous work [15] by combining a genetic introgression approach with whole-genome sequencing to further elucidate the genetic basis of field-derived multiple-anthelmintic resistance in *Teladorsagia circumcincta*, the most economically important nematode pathogen affecting sheep and goats in temperate regions of the world. *T. circumcincta* is a monoxenous, obligately sexual species that infects the fourth stomach (abomasum) of small ruminants, leading to reduced wool, milk and meat production, and in severe cases, death. Widespread anthelmintic resistance has arisen in this trichostrongylid parasite, including multiple-anthelmintic resistance to all major broad-spectrum drug classes available prior to 2008 (i.e., benzimidazoles, imidazothiazoles and macrocyclic lactones, which target microtubule polymerization, nicotinic acetylcholine receptors and glutamate-gated chloride channels respectively) [16] and also to the more recently released amino-acetonitrile derivatives [17]. Through controlled genetic crosses set up by surgical transplantation, we undertook a serial backcrossing experiment that aimed to introgress the resistance-related genes from a field isolate into the genomic background of a partially inbred susceptible recurrent parental strain. Using this partially inbred susceptible parental strain, we generated a draft reference genome of *T. circumcincta* by de novo assembly, which was subsequently used to conduct comparative genome-wide single nucleotide and copy number variant analyses of the resistant strain. In addition, using a combination of pooled (Pool-seq) and individual (ddRAD-seq) genome sequencing and RNA-seq, we identified genes with differential patterns of diversity associated with multiple-anthelmintic resistance.

**Results and discussion**

**The reference genome sequence of *T. circumcincta***

We generated a draft genome of *T. circumcincta* using the partially inbred anthelmintic susceptible strain (S_inbred) which was used as the recurrent parent in the backcrossing program undertaken to introgress anthelmintic resistance-associated genes/alleles into a susceptible genetic background (Fig 1A and 1B). The draft nuclear genome of ~701 Mb (93.4% CEGMA completeness [18]) comprises 81,730 supercontigs (S1 Table), with 35.0% (28,621) of the supercontigs accounting for 90% of the genome. The GC content was 44.8%. The amino acid composition was comparable to that of other phylogenetically close parasitic species such as *Necator americanus* or non-parasitic *Caenorhabditis elegans* (S2 Table). In total, 1,583 repeat families were predicted and annotated, spanning 38.5% of the genome (S3 Table). We predicted a total of 25,532 protein-encoding genes, representing 2.3% of the genome at an average density of 36.4 genes per Mb with an average GC content of 47.8%. Compared to *C. elegans*, the gene density in *T. circumcincta* is lower and the average size of gene loci is larger (Mann–Whitney U test, *P* < 2.2 × 10⁻¹⁶) with longer introns (S1 Table). The majority of predicted genes (80.6%) were supported by transcriptional evidence from mixed-sex adult worm samples with RNA-seq coverage of at least 50% of the length of the annotated coding exons. We predicted secreted proteins (1,603 classical and 9,642 non-classical secretion) and putative membrane-bound proteins (3,749), representing 44% and 15% of the proteome respectively. Functional annotation of deduced proteins on the basis of primary sequence similarity comparisons identified 4,456 unique InterPro domains, 1,563 Gene Ontology terms, and 7,458 KEGG Orthology groups, for 66%, 51%, and 64% of the *T. circumcincta* genes respectively. When considered together, 78% of all *T. circumcincta* genes had some form of putative functional annotation. In spite of our inbreeding (two generations of sibling mating) efforts to reduce heterozygosity in preparation for genome sequencing, the quality of the final assembly...
still suffered from residual heterozygosity, which is consistent with previously reported genome assemblies of obligate outcrossing nematode species with a large effective population size [19, 20]. Although polymorphic haplotypes can be collapsed into consensus sequences during assembly, high genetic diversity tends to result in a fragmented, larger-than-expected assembly [21]. Reliably discriminating uncollapsed alleles from truly paralogous loci remains a significant challenge, and this caveat calls for a careful interpretation of our reference-alignment-based variant analysis.

Introgression mapping of multiple-anthelmintic resistance loci

To analyze genetic variation between the RS³ and S_inbred strains, whole genome re-sequencing analysis was conducted using DNA obtained from pools of 300–500 mixed-sex worms. Approximately 92-fold coverage of the genome was obtained in total across the populations (43.6× and 48.3× from RS³ and S_inbred populations, respectively). Based on the depth of coverage, mapping quality, and gapped regions across all loci, 68.8% of the genome (482.3/700.6 Mb) and 90.6% of the coding sequences (14.9/16.4 Mb) were estimated to have at least the minimum sequence coverage for variant detection in both populations (S4 Table). A set of 17.6 million SNPs was obtained, of which 17.2 million (97.8%) were bi-allelic. The number of segregating (polymorphic) sites was overall ~2-fold lower for the S_inbred strain than for the RS³ strain (7,354,798 vs 16,489,377) (see S5 Table) indicating that the partial inbreeding strategy...
we adopted to reduce heterozygosity in the susceptible reference genome was successful. While a relatively small proportion of SNPs were differentially fixed in the two populations (S_inbred: 4,094; R^3: 147,114 / 17,176,467), there was a notable excess of private SNPs in the RS^3 population (9,617,901 + 19,932 cf 422,182 + 81,072) which were most likely introgressed from the resistant parent strain, R_par. In addition to private R_par derived SNPs, the majority of SNPs observed in the S_inbred population were also segregating in the RS^3 population (6,851,544 / 7,354,798) (S5 Table), as expected from the introgression strategy. Two-dimensional allele frequency spectra based on the bi-allelic sites illustrates this asymmetric distribution of private alleles with concentration of counts in cells along the vertical axis representing the RS^3 population (Fig 1C). The observed pattern is consistent with the expected outcome of our experimental design which relied on a high level of genetic divergence between the two parental isolates (S_inbred and R_par) and a unidirectional gene flow driven by the repeated use of S_inbred in backcrossing. In both populations, low-frequency SNPs were in deficit relative to neutral expectations (genome-wide Tajima’s D: 2.08 and 2.10 for S_inbred and RS^3 respectively), likely due to a combination of ascertainment bias resulting from the limited sampling depth, the exclusion of singleton polymorphisms and the random loss of rare alleles following the inbreeding and introgression strategies employed in the construction of these strains. In addition, the observed level of genetic variability, particularly in the RS^3 population, may be an underestimation considering a possible mapping bias against non-reference alleles. While these biases have the potential to increase uncertainty in population genetic parameter estimation, they likely have limited impact on our ability to detect outlier genomic regions showing the most extreme levels of divergence between the RS^3 and the S_inbred populations.

In the RS^3 parasites we expected the introgressed alleles associated with anthelmintic resistance to be contained in the divergent genomic regions originating from the R_par field isolate, which were maintained by drug selection in the face of repeated gene flow from the S_inbred reference strain. We further expected that independent meiotic recombination events would lead to variation in the introgression break points among the haplotypes segregating in the RS^3 population such that, at the population level, a gradient of allelic divergence would be created peaking around the directly selected loci. A genome-wide scan of F_{ST} following kernel smoothing resulted in demarcation of contiguous regions of the genome with high levels of population differentiation, representing putative introgression blocks. Outlier regions were determined on the basis of the empirical distribution of the smoothed F_{ST} values (Fig 2A) with the goal of prioritizing candidate variants under anthelmintic selection as specific targets for future functional studies. Using 4.5 standard deviations above the mean F_{ST} as a cutoff (i.e., z-score > 4.5; see Methods), genomic regions of ~0.86 Mb were identified across 34 contigs, of which 25 overlapped with a total of 58 protein-coding genes (Fig 2B). Considering the fragmented nature of our draft reference genome, these regions may not all represent independent unlinked loci, particularly when outlier windows are located near the ends of the contigs and thereby miss flanking regions of low F_{ST}. One important consequence of this relatively fragmented assembly is that we cannot be certain of exactly how many high F_{ST} outlier regions (or QTL) differentiate S_inbred and RS^3, so we have focused on high F_{ST} SNPs that fall within these outlier regions, especially where those SNPs fall within or close to predicted genes for which a plausible case can be made for variation in or around that gene to contribute to variation in drug response. Although the size of the individual outlier regions and the number of genes annotated in each were heterogeneous, the majority of the identified regions spanned less than 100 kb (median: 35.5 kb; interquartile range: 27 kb) and harbored less than 4 genes (Fig 2A; S6 Table). Of notable exception was the 290 kb region located on Contig53, which contained 16 outlier genes. While this region may harbor multiple, spatially separated causal variants collectively resisting the gene flow from the S_inbred Population, it is more likely that recombination
has not yet substantially eliminated hitchhiking loci due to a reduced local recombination rate and/or an overall insufficient number of serial backcrosses and drug screening. Although regions containing resistance loci are expected to display higher population differentiation relative to the genomic background, non-uniform distribution of shared ancestral polymorphisms and within-population allelic diversity has likely added a layer of noise to our FST-based introgression mapping approach. At the most fundamental level, however, mapping resolution is limited by the extent to which causal variants are decoupled from neutral hitchhiking loci, and therefore, additional rounds of backcrosses and drug screening would be expected to have helped more fully resolve causal variants from those that are closely linked. Notwithstanding these caveats, this analysis reveals an architecture of resistance genetics that is characterized by multiple regions of elevated [outlier] FST between S\textsubscript{inbred} and RS\textsuperscript{3}. This observation leads to the conclusion that multidrug anthelmintic resistance is likely a polygenic
trait and to the hypothesis that these outlier regions of elevated $F_{ST}$ represent quantitative trait loci (QTL) that are the products of selection for resistance.

To test this hypothesis, and to help prioritize the candidate genes located in the outlier regions (Fig 2; S7 Table) for more detailed analysis, we critically evaluated whether any of these genes have known/predicted functions that can be plausibly connected to anthelmintic resistance in light of our current understanding of the mechanisms of drug action [7]. Among the most notable candidates were a β-tubulin gene (TELCIR_01271), a major target of benzimidazole anthelmintics, and putative orthologs of *Cel-unc-29* nicotine acetylcholine receptor (nAChR) subunit (TELCIR_06180) that may constitute a component of a levamisole-sensitive receptor in *T. circumcincta*. Interestingly, additional members of the Cys-loop ligand-gated ion channel families (LGICs) were represented including putative orthologs of *Cel-acr-11* nAChR (TELCIR_03607) and the *Cel-lgc-54* LGIC (TELCIR_00170). Overall, GO terms attributable to these LGICs, such as acetylcholine-activated cation-selective channel activity ($P = 2.6 \times 10^{-3}$), extracellular ligand-gated ion channel activity ($P = 6.3 \times 10^{-3}$), and postsynaptic membrane ($P = 6.7 \times 10^{-3}$) were significantly overrepresented among the outlier genes (Table 1). These results underscore the potential contribution of target gene variation in the development of anthelmintic resistance in *T. circumcincta* under field conditions, although our data do not rule out a more complex genetic architecture involving additional, presently uncharacterized genes. To further contextualize our findings in relation to previously reported anthelmintic resistance-associated genes and gene families, we examined gene-wise $F_{ST}$ values and highlighted non-synonymous SNPs that showed not only significant ($P < 1.0 \times 10^{-5}$) but indeed substantial differentiation between the susceptible and the resistant strains (Fig 3; S7 and S8 Tables).

In the present genome assembly, we identified two paralogs of β-tubulin (*isotype-1* and -2) that are co-orthologous to *Cel-ben-1*, the locus which confers benzimidazole (BZ) sensitivity in *C. elegans* [22]. This finding is in line with the model of a lineage-specific duplication in trichostrongylid species [23]. Both isotypes have been implicated in BZ resistance [24, 25], with the hypothesis that selection occurs in two stages [26]: an initial reduction in diversity at *isotype-1* followed by the loss of *isotype-2*. We observed only *isotype-1* (TELCIR_01271) variation associated with the outlier loci with high $F_{ST}$ (z-score > 4.5) in our genome-wide survey of the resistant backcross progeny (RS3). Furthermore, no evidence of selection was detected in any of the remaining members of the *T. circumcincta* β-tubulin gene family (Fig 3A). In the case of β-tubulin *isotype-1*, two non-synonymous coding variants, E198L (GAa/TTa) and F200Y (tTc/tAc), were exclusively found in the RS3 population and present at allele frequencies of 28.1% and 72.8%, respectively. The F200Y variant confers BZ resistance in *H. contortus* [27], and has been widely recognized in many species of parasitic nematode as a major resistance determinant. Although amino acid substitutions at position 198 (e.g., E198A) are less common in nematodes, variants at this position have been linked to BZ resistance phenotypes in *H.

Table 1. Over-represented GO terms among genes located in $F_{ST}$ outlier regions (z-score > 4.5).

<table>
<thead>
<tr>
<th>Gene Ontology</th>
<th>Hypergeometric test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P-value</td>
</tr>
<tr>
<td>Cellular component</td>
<td></td>
</tr>
<tr>
<td>Postsynaptic membrane</td>
<td>6.7 × 10^{-3}</td>
</tr>
<tr>
<td>Synaptic part</td>
<td>9.8 × 10^{-3}</td>
</tr>
<tr>
<td>Molecular function</td>
<td></td>
</tr>
<tr>
<td>Acetylcholine-activated cation-selective channel activity</td>
<td>2.6 × 10^{-3}</td>
</tr>
<tr>
<td>Extracellular ligand-gated ion channel activity</td>
<td>6.3 × 10^{-3}</td>
</tr>
<tr>
<td>Ligand-gated channel activity</td>
<td>7.0 × 10^{-3}</td>
</tr>
<tr>
<td>DNA-directed RNA polymerase activity</td>
<td>8.4 × 10^{-3}</td>
</tr>
<tr>
<td>Ion channel activity</td>
<td>8.8 × 10^{-3}</td>
</tr>
<tr>
<td>Channel activity</td>
<td>8.8 × 10^{-3}</td>
</tr>
</tbody>
</table>

[10.1371/journal.pgen.1006857.t001](https://doi.org/10.1371/journal.pgen.1006857.t001)
Non-synonymous coding variants identified in the *Saccharomyces cerevisiae* strain revealed haplotypes harboring E198L (GAa/TTa) and F200Y (tTc/tAc) variants. RS mechanism [30]. We reconstructed three segregating haplotypes of *R. stricte~, especially worms heterozygous for only P198, was consistent with (and also sup-

The inferred haplotype structure indicates that (i) F200Y occurs on at least two distinct and diverse haplotype backgrounds, suggesting multiple independent origins of the variant allele, and (ii) E198L and F200Y variants occur in *trans* on separate haplotypes. In agreement with this haplotype reconstruction, genotyping of individual male worms from the S population revealed haplotypes harboring E198L and F200Y (TTc/tAc) variants. 

https://doi.org/10.1371/journal.pgen.1006857.g003

200

β-tubulins, (B) ligand-gated cation channels, (C) ligand-gated anion channels, and (D) ATP-binding cassette transporters in *Teladorsagia circumcincta*. Unsupported nodes (bootstrap support less than 50%) were collapsed to polytomy. The shading on the trees highlights monophyletic groups. *Caenorhabditis elegans* (Ce) and *Haemonchus contortus* (Hco) homologs were included to help resolve the phylogeny. Tubulin from *Saccharomyces cerevisiae* (Sce-TUB-2) served as an outgroup. Genome-wide fixation index (FST) and copy number variation (CNV) between the RS3 and S inbred populations of *T. circumcincta* were represented as a heatmap. Non-synonymous coding variants identified in the FST outlier genes (z-score > 4.5) were reported and loci with >50% allele frequency differences were indicated in bold. *De novo* assembly [101] of β-tubulin isotype-1 (TELCIR_01271) from the RS3 strain revealed haplotypes harboring E198L (GAa/TTa) and F200Y (TTc/tAc) variants. 

Introgression mapping of anthelmintic resistance

contortus* [28] and, more recently, in *T. circumcincta* [29], and molecular modeling suggests that the associated loss of hydrogen bonding interactions may play a role in the resistance mechanism [30]. We reconstructed three segregating haplotypes of β-tubulin *isotype-1* in the RS3 population over the exonic region harboring E198L and F200Y variants (Fig 3A; S1 Fig). The inferred haplotype structure indicates that (i) F200Y occurs on at least two distinct and diverse haplotype backgrounds, suggesting multiple independent origins of the variant allele, and (ii) E198L and F200Y variants occur in *trans* on separate haplotypes. In agreement with this haplotype reconstruction, genotyping of individual male worms from the S inbred (n = 94) and RS3 (n = 79) populations failed to detect any individuals homozygous for both resistance alleles (R198/R198/R200/R200) although worms homozygous resistant at one locus only (S198S198/R200 R200 and R198R198/S200/S200) were observed, as were double heterozygotes (S198R198/ S200R200) (S2 Fig; S9 Table). Considering that S198R200 haplotype was segregating in the RS3 population at a minimum inferred frequency of 79.7% (S9 Table), the absence of single heterozygotes (especially worms homozygous for only P198) was consistent with (and also supported) the conclusion of our Pool-seq analysis that R198R200 haplotype was not present in the RS3 population. Sequencing of *isotype-1* from the same individuals confirmed the existence of multiple haplotypes, supporting the conclusion that BZ resistance conferring alleles arose several times in the OR parent of RS3 (S3 Fig), as has been reported for BZ resistance conferring alleles in the UK [29]. Although further work will be necessary to fully determine the extent to which E198L contributes to the overall resistance phenotype, the absence of haplotype(s) simultaneously harboring both variants suggests that, under field conditions, E198L can confer BZ resistance (and hence was selected) independently of F200Y, and that β-tubulin carrying both variants either is detrimental to organismal fitness (i.e., negative intramolecular epistasis) or the double mutation event (or recombination over an interval of <6bp) required to give rise to a *cis* haplotype is sufficiently unlikely that it is not observed.

Cholinergic anthelmintics, such as LEV and pyrantel, induce spastic neuromuscular paralysis by selectively opening nAChRs, a family of pentamic ion channels belonging to the Cys-loop LGIC superfamily that convert neurotransmitter binding into membrane electrical depolarization. Each receptor subunit has an N-terminal extracellular ligand-binding domain (ECD) followed by four transmembrane helices (TMD) that form the ion channel [31]. Nematode genomes encode a large number of nAChR subunits (~30) [32] and different subunit combinations result in pharmacologically distinct receptors [33]. Although the precise subunit composition of LEV-sensitive nAChR in *T. circumcincta* has not yet been determined, a putative model for trichostrongylid species suggests a likely involvement of parasite orthologs of *Cel-unc-29*, *Cel-unc-38*, *Cel-unc-64* and *Cel-acr-8* in receptor formation [33], and mutation, truncation and decreased expression of these subunits have been observed in field-selected LEV-resistant trichostrongylids [34–36]. *Tci-unc-29.4* and *Tci-acr-11* nAChR genes were identified in our FST outlier analysis and, in addition, the genomic regions encoding *Tci-unc-29.2* (TELCIR_06181) and *Tci-acr-8* (TELCIR_04316), displayed relatively high FST values (z-score of 4.32 and 4.18, respectively) (Fig 3B). Within these subunit genes, 16 non-synonymous coding
variants were found. When considering the potential consequences of the substitutions based on the amino acid properties [37] and their locations relative to the ligand-binding or the transmembrane domains, the probability that any of these variants have drastic detrimental effects on protein structure and function appears to be relatively low. This observation is consistent with the view that, under field conditions, loss of function variants are likely to experience negative selection due to reduced fitness. Several of the variants (e.g., T388I and '462W (stop loss) in Tci-acr-11 and P24Q in Tci-acr-8) appear to have a greater potential to alter protein function, although the allele frequency of these variants in the resistant population is not substantially different from that in the susceptible population, suggesting that they are unlikely to play a direct role in LEV resistance (S8 Table). Further biochemical, pharmacological and structural modeling work will be necessary to fully assess and understand the impact of these alleles on drug-target interactions. These results also raise the possibility that drug selection may have acted primarily on the noncoding regulatory variants of the FST outlier nAChR genes. A preliminary assessment of the transcript abundance levels (see Methods) suggested that Tci-acr-11 transcript was substantially less abundant in the resistant strain (RS³) relative to the susceptible strain (S³ inbred) (RS³/S³ inbred log2 ratio: -5.02; see S7 Table), in a manner similar to previous reports that showed decreased expression of nAChR subunit genes in various LEV-resistant trichostrongylid populations [38–40]. However, because of the limited mapping resolution of the present study and our generally poor understanding of the functional consequences of noncoding variants, we are unable to identify specific candidate noncoding mutation(s) that could contribute to the resistance phenotype. Furthermore, although mutant screens in C. elegans indicate that genetic variations in calcium-mediated muscle contraction signaling pathway and ancillary proteins involved in nAChR assembly/maintenance may influence LEV susceptibility [41, 42], we did not observe any significant evidence of genetic differentiation among the genes implicated in the LEV excitation-contraction pathway in the RS³ population (S7 Table).

The predicted gene TELCIR_00170 (Tci-lgc-54) is one of the top FST outliers in our analysis (S7 Table). It belongs to the Cys-loop ligand gated chloride channel branch of the LGIC superfamily and is distinct from the nAChRs associated with LEV resistance (Fig 3C) and from GluCl and GABA receptor family members. The likely C. elegans ortholog, Cel-lgc-54, is described as a predicted “ligand unknown” biogenic amine-gated chloride channel [43] and as a GABA-receptor [44] but has not been implicated previously in relation to IVM resistance. Although the ligand for nematode LGC-54 is not yet identified, the predicted protein contains a tryptophan in ligand-binding loop C (amino acid position 231), which has been hypothesized to be a key residue for binding amines [45], and it is known that other family members (including Cel-lgc-55, the most closely related paralog in C. elegans) are activated by serotonin, dopamine and tyramine [46–48]. Furthermore, a gene encoding a putative dopamine receptor, Hco-ggr-3, has been implicated in IVM resistance in H. contortus [48]. In the RS³ population, we identified 5 non-synonymous variants in Tci-lgc-54: T9S, K11E, S18fs, A20P, and S336N. The former four are located upstream of the N-terminal ECD, and the latter is located at the beginning of the cytosolic loop between the third (M3) and the fourth (M4) TMD alpha helices. Notably, the frameshift variant at position 18 introduces a premature stop codon and the A20P variant located at the predicted signal peptide cleavage site (position 20–21) has a potential to interfere with the proper signal peptide processing (S4 Fig). Failure of signal peptide cleavage is likely to result in mislocation and/or degradation of the protein and thus behave as a loss-of-function mutation. It is of interest in this context that (a) large deletion alleles of Cel-lgc-54 and Cel-ggr-3 in C. elegans are viable suggesting that these channel subunits are not essential (although loss-of-function mutations in Cel-lgc-55 confer subtle behavioral phenotypes [47]), (b) reduced IVM sensitivity in H. contortus is associated with reduction in the
transcript abundance of Hco-ggr-3 [48] and (c) a four-amino-acid deletion in the N-terminal region of Cel-gl-1 has been linked to IVM resistance in C. elegans [49].

Although glutamate-gated chloride channels (GluCls) are considered its main targets, IVM may also interact directly with other anionic Cys-loop LGICs, including GABA_A and glycine receptors [50, 51] and irreversible activation of these inhibitory chloride channels by IVM results in flaccid paralysis and eventual expulsion of the parasite [52, 53]. Several glutamate and GABA gated chloride channel genes have been implicated in IVM resistance in H. contortus (e.g. Beech et al. 2013) but none of these genes showed significant values for FST in our analysis of T. circumcincta. In addition to genes involved directly or indirectly in neurotransmitter functions, genes putatively responsible for amphid neuron defects in C. elegans and H. contortus, such as Cel-che-3, Cel-dyf-7 and Hco-dyf-7, have been implicated in IVM resistance in these species [54–56]. Again, none of the likely che-3 and dyf-7 orthologs in T. circumcincta displayed notably high FST values in our analysis (S7 Table). Although it is possible that different IVM resistance mechanisms are involved in different nematode species, a recent study of UK field populations of H. contortus, no evidence of selection by IVM was detected for Hco-lgc-37, Hco-glc-5, Hco-avr-14 or Hco-dyf-7 [57]. Furthermore, a recent analysis of IVM resistant backcross populations in H. contortus has also suggested that these candidate genes were not associated with resistance [58]. It is thus conceivable that some of the putative candidate genes from earlier single-locus studies may represent false-positive associations. If, as we conclude on the basis of the data reported here, multidrug anthelmintic resistance is a polygenic quantitative trait, one explanation for this apparent discrepancy is that in the resistant population that we analyzed these genes are not under selection. The observation that >1 genotype can result in the same phenotype is expected for quantitative traits. This may also, for example, explain why macrocyclic lactone resistance in particular seems so genetically heterogeneous, with many different candidates apparently under selection in different resistant populations.

Copy number variations associated with multiple-anthelmintic resistance

We also examined copy number variations (CNVs) between the RS^3 and S_inbred strains of T. circumcincta (Fig 2C). Since our reference assembly was generated using the S_inbred strain, we focused our analysis on genomic regions displaying increased copy number in the RS^3 population relative to the S_inbred Population, because any copy number decrease in the RS^3 strain was likely to have been confounded with a potential mapping bias against highly divergent reads.

Table 2. Top 10 CNV regions that contain protein-coding genes in Teladorsagia circumcincta RS^3 strain showing the highest level of copy number increase relative to the S_inbred strain.

<table>
<thead>
<tr>
<th>CNV region</th>
<th>Log2 ratio (RS^3/S_inbred)</th>
<th>Overlapping gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cont17796:492–5412</td>
<td>8.8</td>
<td>TELCIR_20472</td>
</tr>
<tr>
<td>Cont4607:20664–27060</td>
<td>7.2</td>
<td>TELCIR_13581</td>
</tr>
<tr>
<td>Cont13316:492–8364</td>
<td>6.1</td>
<td>TELCIR_19247</td>
</tr>
<tr>
<td>Cont19663:492–2460</td>
<td>6.0</td>
<td>TELCIR_23120</td>
</tr>
<tr>
<td>Cont15521:492–3936</td>
<td>4.8</td>
<td>TELCIR_19884</td>
</tr>
<tr>
<td>Cont5:324720–326196</td>
<td>4.4</td>
<td>TELCIR_00177</td>
</tr>
<tr>
<td>Cont204265:492–1968</td>
<td>3.3</td>
<td>TELCIR_23838</td>
</tr>
<tr>
<td>Cont4815:28044–30504</td>
<td>3.3</td>
<td>TELCIR_13812</td>
</tr>
<tr>
<td>Cont1731:492–7872</td>
<td>2.9</td>
<td>TELCIR_08884</td>
</tr>
<tr>
<td>Cont1090:71832–80196</td>
<td>2.9</td>
<td>TELCIR_07060</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pgen.1006857.t002
containing non-reference alleles (especially in the intronic and intergenic regions). The top 10 protein-coding CNV regions showing the most extreme inter-strain variation (log$_2$ read count ratio > 2.9) contained 13 genes, 4 of which are likely orthologs of *C. elegans* P-glycoprotein 9 (*Tci-pgp-9; TELCIR_08885, TELCIR_13813, TELCIR_19247 and TELCIR_19884*) (Table 2; Fig 3D). These sequences (one complete and three partial genes) appear to represent the individual haplotypes of *Tci-pgp-9* that are segregating in the *S$_{inbred}$* population. While our pooled sequencing data provide strong evidence of an increase in *Tci-pgp-9* copy number on average in the resistant population relative to the susceptible population, it remains challenging to reliably resolve the full haplotype sequences and their respective within-population copy number variability for each population. Nevertheless, confirmation that RS$^3$ strain parasites carry additional copies of this gene compared to *S$_{inbred}$* parasites was provided by a separate investigation of *Tci-pgp-9* copy number using single worm genomic DNA quantitative PCR (S5 Fig). Furthermore, these single worm data showed that certain *Tci-pgp-9* haplotypes (i.e., allelic variants defined on the basis of the first inter-nucleotide binding domain (IBDA) sequence polymorphisms) (S6 Fig) occurred only in worms exhibiting increased *Tci-pgp-9* copy number (IBDA haplotypes 3, 6 and 10) (S10 Table), suggesting (a) that these haplotypes arose as a result of the gene duplication event(s) that gave rise to the increase in copy number, (b) that the duplication(s) occurred long enough ago that the duplicated copies have started to diverge and/or (c) selection for or against functional differences in IVM-affinity conferred by specific haplotypes. One further haplotype that appeared to be enriched in the RS$^3$ population, haplotype 2, does not occur in worms that carry additional copies of *Tci-pgp-9* (S10 Table), suggesting that selection for this haplotype in resistant worms is not related to increased copy number.

P-glycoprotein (P-gp) is an ATP-binding cassette (ABC) transporter with two homologous halves, each containing a TMD and a cytoplasmic nucleotide-binding domain (NBD). Using the energy from ATP hydrolysis, P-gp actively transports many lipophilic compounds (both endogenous metabolites and xenobiotics) out of the cell from the inner leaflet of the membrane, providing a mechanism by which anthelmintic concentration at the receptor site may be reduced. Sequence polymorphism and constitutive or inducible overexpression of P-gp’s have been reported in IVM-resistant populations of several nematode species, including *T. circumcincta*, in support of the hypothesis that an increased drug efflux due to changes in expression, activity and/or substrate specificity of ABC transporters can contribute to IVM resistance [59, 60]. Our preliminary assessment (by RNA-seq) indicates that *Tci-pgp-9* transcripts are more abundant in the RS$^3$ population relative to the *S$_{inbred}$* population (log$_2$ ratio > 3) (S7 Table), suggesting that the *Tci-pgp-9* copy number increase facilitates (constitutive or inducible) increased expression of the transporter in the resistant population. Cloning and sequencing of individual cDNA transcripts corresponding to the N- and C-terminal transmembrane domains (with their extracellular loops) showed that the predominant transcripts in the RS$^3$ population carry either a splice variant that results in a deletion of 45 aa from the first predicted extracellular loop between TM1 and TM2 (a region of the protein hypothesized to play a role in substrate binding), or a full length variant that contains 3 non-synonymous amino acid substitutions in the same predicted loop (S7–S9 Figs). Thus, it appears likely that a combination of increased expression (via increased copy number) and sequence polymorphism may contribute to the association between *Tci-pgp-9* and IVM resistance in the RS$^3$ strain.

Mapping of genetic loci associated with ivermectin resistance based on single worm genotyping

The design of our introgression strategy, which aimed to identify genomic regions concurrently selected in response to BZ, LEV and IVM, did not allow us to directly assess the relative
contribution of candidate loci to resistance against each of the individual anthelmintic classes. We therefore undertook a complementary mapping approach with a specific focus on IVM resistance using F2 populations derived from a cross between the S_{inh}red susceptible and the R_{par} multiple-anthelmintic resistant isolates, i.e., the parental strains of the RS^{3} backcross progeny population on which our whole-genome introgression study was based (Fig 1A). Individuals from IVM-screened and drug-naïve F2 mapping populations (n = 24 male worms from each group) were genotyped by ddRAD-seq, a reduced-representation genome sequencing method, yielding a total of 0.59 million variant calls. Using the \( F_{ST} \) estimates of segregating SNPs that satisfied a minimum sampling depth of 10 individuals in both populations (n = 2,628) (S11 Table), we identified outlier loci and linked genes (i.e., contigs) most strongly differentiated in the IVM-survivor group relative to the drug-naïve control group, and compared the outcome against the outlier genes identified from the introgression mapping experiment. Even though there was a high rate of allele dropout (most likely due to sequence polymorphism within the restriction sites in our mapping population leading to the loss of affected restriction fragments from our RAD-seq libraries), we were able to survey part of the genome (349 contigs; combined length = 57.6 Mb or ~8% of the genome) for outlier loci. It has been shown that, in RAD-seq experiments, missing data can inflate \( F_{ST} \) values and rates of false-positive outliers increase as the chromosome sampling depth cutoff decreases [61]. We indeed observed a strong dependency between the variance of \( F_{ST} \) estimates and the allele dropout (S10 Fig) and therefore determined outliers in each individual sampling depth category separately under the assumption that outliers were evenly distributed across loci irrespective of missing data (Fig 2D).

Within the part of the genome that was subjected to \( F_{ST} \) outlier analysis, we identified 18 genes across 5 contigs that displayed some evidence of genetic differentiation in both the introgression and F2 mapping experiments (i.e., minimum \( F_{ST} \) z-score of 2.5 in both datasets) (S12 Table). Included in this combined list is \( Tci-lgc-54 \), one of the top outlier genes from the genomic introgression analysis. Although the evidence of selection is not as strong in the IVM-survivor F2 mapping population as it was in the introgressed multiple-anthelmintic resistant population (RS^{3}), the amine-gated chloride channel \( Tci-lgc-54 \) is the only candidate LGIC that is supported strongly by both of the mapping approaches (S13 Table) and thus merits further analysis as a potential IVM resistance gene in \( T. circumcincta \). It is important to note that the individual F2 generation IVM treatment survivors analyzed here were not significantly more resistant to either BZ or LEV treatment as a result of the IVM treatment (see S14 and S15 Tables). Thus the resistance phenotypes appeared to have segregated independently, indicating that \( Tci-lgc-54 \) was an \( F_{ST} \) outlier in IVM-resistant F2 segregants that remained susceptible to BZ and LEV treatment. Consequently, despite strong evidence for selection of ABC transporters, these data do not support a single “multidrug resistance mechanism” able to confer resistance simultaneously to all 3 drug classes.

We were unable to assess whether the \( Tci-pgp-9 \) revealed by Introgression analysis as a strong candidate IVM-resistance locus co-segregates with IVM resistance in the F2 mapping population because of the absence in the RAD-seq data of linked SNP makers with adequate sampling depth for \( F_{ST} \) outlier analysis. A different ABC transporter, \( Tci-mrp-6 \) (TEL-CIR_03131) (Fig 3D), is however present on the combined introgression/segregation list. Our results suggest the interesting possibility that multiple ABC transporters may be involved in IVM efflux, either within the same worm or in different worms, in the RS^{3} multiple-anthelmintic resistant population. In support of this conclusion, several reports in other parasitic nematodes implicate a range of ABC transporter family members [62–64], implying that there may be many combinations of ABC transporters able to contribute to IVM-resistance.

A novel, strongly-supported, candidate region in our combined outlier list (S12 Table) is on Contig209, which contains two putative triacylglycerol lipase genes (TEL-CIR_02985 and...
TELCIR_02988; FST z-score > 3.3). Intriguingly, a triacylglycerol lipase/cholesterol esterase gene (F54F3.3) has been shown in *C. elegans* to respond transcriptionally to IVM exposure [65], although it remains to be determined whether the lipase activity plays a role in a drug-induced starvation-related stress response that facilitates tolerance of or recovery from ivermectin toxicity, or whether lipid metabolism may play a more direct role in IVM metabolism or detoxification. It is of interest to note that a similar pool sequencing analysis of ivermectin response in *Onchocerca volvulus* [66] points to the involvement of likely orthologues of these genes in a distantly related nematode parasite of significant medical importance.

The work reported here is the first to combine classical genetic methods such as introgression and segregation analysis with new genomic tools such as RAD-seq and whole genome re-sequencing to analyze multiple-antihelminthic resistance in a parasitic nematode. The nematode species examined, *T. circumcincta*, is an economically significant, globally distributed gastrointestinal parasite of small ruminants. More importantly however, with the rapid proliferation of mass drug administration programs globally for treatment of helminth infections of humans, there is an urgent need to better understand the genetic basis of resistance to the drugs that form the basis of those programs. We show clearly that resistance to each of the three drug classes segregates independently of the others and that for LEV and IVM resistance in particular, multiple loci likely contribute to the resistance in a variety of ways (possible reduction/modulation in target site sensitivity, reduced target site expression, increased drug efflux, etc.), so that drug resistance in these parasites should best be thought of as a multifactorial quantitative trait rather than a simple, discrete Mendelian character. The polygenic genetic architecture of resistance provides an explanation for the apparent discrepancies between the many single, candidate gene studies and, since many genes can contribute to resistance, it seems likely therefore that the combination selected in any given circumstance is as likely to be a product of genetic drift as of selection *per se*. Furthermore, it is also clear from these data that alleles that contribute to resistance for each drug class have arisen many times on different genetic backgrounds, giving rise to a heterogeneous mix of “resistance haplotypes” that implies soft rather than hard selective sweeps. This is most obvious at the β-tubulin *isotype-1* locus. Although selection at this locus appears to be necessary and sufficient for BZ-resistance, we observed extensive polymorphism surrounding the amino acid 198 and 200 determinants of resistance, suggesting soft selection from multiple pre-existing variants at P198 or P200 rather than hard selection of a single allele at a single position has occurred for this resistance. Similar genetic heterogeneity at this locus has been observed in other BZ-resistant isolates of *T. circumcincta* and *H. contortus* [29].

In conclusion, this work elucidates the polygenic, quantitative trait genetic architecture of multiple anthelmintic resistance in a parasitic nematode for the first time and establishes a framework on which future studies of the inevitable evolution of anthelmintic resistance in nematode parasites of humans can be based. In this context, it is significant that a similar study of ivermectin response in *O. volvulus* [66] points to a similar genetic architecture, with hits either to likely orthologues of genes identified here or to similar neuronal functions, thus demonstrating the utility of studies in more tractable parasite species such as *T. circumcincta*.

**Materials and methods**

**Ethics statement**

All experimental procedures used in generating the parasite material for this study were approved by AgResearch’s Wallaceville Animal Research Centre Animal Ethics Committee under the Animal Welfare Act 1999 in New Zealand [AEC application numbers 516, 562 & 636].
Parental isolates used in developing the introgressed strain

The multiple-anthelmintic resistant field strain of *T. circumcincta* (OR strain) used in this study was isolated in New Zealand in 1996 from lambs which had been grazing a property previously occupied by Angora goats (Leathwick DM, personal communication). Fecal nematode egg count reduction tests undertaken on the lambs revealed that none of the three broad-spectrum anthelmintic families available at that time, i.e., oxendazole (BZ), levamisole (LEV) and ivermectin (IVM), were fully effective against this isolate. Prior to inter-strain crosses being set up, a population of OR was maintained for five generations in pen-raised goat kids, and screened at each generation with selected representatives of BZ (Systamex; Schering Plough, Kenilworth, NJ; 4.5 mg/kg), LEV (Levicare; Ancare New Zealand, Auckland, New Zealand; 7.5 mg/kg) and IVM (Ivomec liquid for sheep and goats, Merial New Zealand; 0.2 mg/kg) to maximize the proportion of worms homozygous for resistance to each of them. Anthelmintics were administered to the goat kids at the manufacturer’s recommended dose rate unless no specific goat dose rate was provided, in which case a standard sheep dose rate was used (as was common practice on goat farms in New Zealand before the widespread emergence of anthelmintic resistance in this host species). The efficacy of all these drugs, used either individually or in combination, was very low against the resulting resistant parental (R<sub>par</sub>) strain. The anthelmintic susceptible S strain of *T. circumcincta* was originally isolated from field-grazed lambs in New Zealand during the 1950’s prior to the widespread use of broad-spectrum anthelmintics (Elliott DC, personal communication). This isolate had subsequently been maintained at Wallaceville Animal Research Centre (AgResearch, New Zealand) by annual passage through pen-raised drug-naïve lambs. Given that significant genetic diversity can be maintained even in laboratory-passaged nematode populations of limited size, the S isolate was subjected to two generations of half-sib mating in an attempt to reduce the background genetic variance in preparation for introgression mapping. Briefly, mature eggs were collected from the oviduct of a single gravid adult female recovered from the host’s abomasum, and cultured to collect infective larvae. Thirteen of these sibling larvae were used to orally infect a pen-raised parasite-free goat kid, and half of the resulting progeny were used to re-infect the same host to supplement the existing infection. A second goat kid was subsequently infected using larvae cultured from the first. Sibling mating was repeated as before by isolating and culturing eggs from a single adult female worm isolated from the abomasum of the second kid. Twenty sibling larvae from this culture were used to infect a third worm-free goat kid whose fecal output was cultured for subsequent infections. Anthelmintic efficacy testing on this partially inbred S strain (S<sub>inbred</sub>) revealed that representatives of all three broad-spectrum anthelmintic classes were highly effective (efficacy >99.0% for BZ, LEV, and IVM) [15].

Introgression of multiple-anthelmintic resistance genes into an inbred anthelmintic susceptible genetic background by serial backcrossing

A schematic of the backcrossing and selection experiment is outlined in Fig 1A. Crosses between R<sub>par</sub> and S<sub>inbred</sub> strains of *T. circumcincta* were performed by surgical transfer of worms from separate donor goat kids (containing either R<sub>par</sub> or S<sub>inbred</sub> worms) into the abomasum of a recipient goat kid. In order to ensure that the female S<sub>inbred</sub> worms had not yet mated, infection of the donor kids was timed so that the females would be 10 days old and thus still at the late-fourth developmental stage (L<sub>4</sub>) at the time of transfer, while male R<sub>par</sub> worms would be 5 weeks old and thus adults. Ten days before transfer, the R<sub>par</sub> worms were screened with BZ, LEV and IVM. Worms were collected from donor goat abomasas, rinsed with phosphate-buffered saline over a 45µm Endecott sieve, and inspected microscopically. Approximately 300 male adult male (R<sub>par</sub>) and an equivalent number of L<sub>4</sub> female (S<sub>inbred</sub>) worms were
collected into a modified Nematode Growth Medium containing 0.4% w/v agar [67], and surgically transferred into the abomasum of a previously worm-free recipient goat kid. The [heterozygous] F1 progeny resulting from this cross were then used to infect another worm-free kid to obtain an F2 generation in which the alleles for anthelmintic resistance were expected to have segregated. The F2 infective larvae cultured from this kid were used to infect a further worm-free goat kid, which was then subjected to successive doses of IVM, BZ and LEV over a period of 24 hrs at 28 days post-infection so that the F3 generation would be derived from worms carrying the full complement of genes needed for multiple-anthelmintic resistance. In order to maximize the frequency of resistance alleles in the population, F3 worms were passed for a further two generations of drug screening with IVM, BZ and LEV. At this point a backcross between the anthelmintic-screened F5 generation and Snbred worms was set up using similar procedures to those described for the initial crosses. F2, F3 and F4 generations of the backcross worms were then each screened, as before, with all 3 anthelmintic classes before a final round of backcrossing and a further 4 generations of drug screening. Because each generation of backcrossing reduces the proportion of the donor parent genome present in the population by half, the resultant multiple-anthelmintic resistant worm population (RS3) was expected to have a genetic makeup largely similar (7/8) to that of the susceptible recurrent parent (Snbred) but, at the same time, be carrying the anthelmintic-resistance genes derived from the Rpar strain. Crosses based on mass mating and multiple generations of drug selection (between and after backcrosses) were an important design feature to create variation in recombination breakpoints and divide individual introgression blocks (i.e., segments of DNA of Rpar origin untouched by recombination) into smaller fragments.

**Preparation of genomic DNA and total RNA for high-throughput sequencing**

Parasite-free lambs, maintained indoors on a diet designed to avoid any unintended nematode infections, were each infected with approximately 24,000 larvae of either the RS3 strain (2 lambs) or the Snbred strain (3 lambs) of *T. circumcincta*. At 28 days post-infection, lambs that had received the RS3 strain were treated successively with IVM, BZ and LEV over a period of 24 hrs. At 37 days post-infection (9 days post-treatment), the RS3 worms were collected from the abomasum, washed free of all debris in physiological saline, and then transferred in mixed sex batches of 300–500 individuals into 1.5 ml tubes in which they were snap frozen at -80˚C. The Snbred worms were similarly collected and snap frozen at 28 days post-infection without anthelmintic treatment. Genomic DNA was isolated from each of the strains using a method modified from that described by Sulston and Hodgkin [68]. Each worm sample (after thawing) was suspended in 150–200 μl of lysis solution [100 mM NaCl, 100 mM Tris-HCl (pH 8.5), 50 mM EDTA (pH 7.5), 1% SDS, 2% β-mercaptoethanol and 200 μg/ml proteinase K], placed in a glass/teflon tissue grinder and thoroughly homogenized. The homogenate was transferred into a sterile 10 ml tube and supplemented with additional lysis solution to a total volume of 3 ml. After incubation at 65˚C for 16 hr with gentle mixing, 10 μl of RNase A (100 mg/ml; QIAgen, Hilden, Germany) was added and the lysate incubated for 10 min at 45˚C. To remove protein/poly saccharide complexes (which can be problematic in nematode DNA preparations), 750 μl 5M NaCl and 500 μl CTAB/NaCl (10% CTAB in 0.7M NaCl) were added, and after gentle mixing the tube was incubated for 15 min at 65˚C. The lysate was then extracted successively with equal volumes of phenol/chloroform/isoamyl alcohol (25/24/1) and chloroform/isoamyl alcohol (24/1). Following recovery of the aqueous layer from the final extraction, DNA was precipitated by the addition of 2 volumes of absolute ethanol (4˚C), pelleted by centrifugation, washed twice in 70% ethanol (4˚C), briefly air-dried, and re-suspended in 150 μl 10mM TrisCl (pH 8.5). Total RNA was also prepared from the mixed-sex batches of adult *T. circumcincta*
from each of the S\textsubscript{inbred} and RS\textsuperscript{3} strains. Frozen worm samples were added to 1 ml of TRIzol reagent (Invitrogen, Carlsbad, CA) and the mixture ground to a powder in liquid nitrogen in a mortar and pestle. The resulting powder was transferred into a microfuge tube, 200 μl chloroform was added and the tube was shaken vigorously before being centrifuged at 12,000g for 15 min at 4˚C. Following recovery of the upper phase, the RNA was precipitated by the addition of 500 μl of isopropanol and pelleted by centrifugation at 12,000g for 10 min at 4˚C. The RNA pellet was then washed in 75% ethanol and air-dried briefly before re-suspending in 40 μl UltraPure DNase/RNase Free Distilled Water (Invitrogen). The integrity and yield of the RNA was verified using the Bioanalyzer 2100 (Agilent Technologies, Cedar Creek, TX).

**Sequencing, de novo assembly and annotation of an anthelmintic-susceptible T. circumcincta reference genome**

As a basis for introgression mapping and comprehensive variant analysis, we generated a draft genome sequence for the anthelmintic-susceptible S\textsubscript{inbred} strain of T. circumcincta (BioProject ID: PRJNA72569). Whole genome shotgun libraries (fragments and mean insert size of 3kb and 8kb) were generated as previously described [69] and sequenced using a Genome Sequencer Titanium FLX (Roche Diagnostics, Basel, Switzerland) platform (S16 Table), and assembled using Newbler v. 2.6 [70]. To improve scaffolding, an in-house tool CIGA (Cdna tool for Improving Genome Assembly) was used to map 454 cDNA reads using BLAT [71] to the genomic assembly to link genomic contigs. Gaps were then closed using Pygap, an in-house tool, which utilizes the Pyramid assembler and uses Illumina paired-end reads to close gaps and extend contigs. The repeat library was generated using Repeatmodeler (http://repeatmasker.org), and Tandem Repeat Finder [72] was used in addition for sequence annotation. Repeats and predicted RNAs were then masked using RepeatMasker (http://repeatmasker.org). The ribosomal RNA genes were identified using RNAmmer [73] and transfer RNAs were identified with tRNAscan-SE [74]. Non-coding RNAs, such as microRNAs, were identified by sequence homology search of the Rfam database [75]. Protein-coding genes were predicted using a combination of \textit{ab initio} programs Snap [76], Fgenesh [77] and Augustus [78] and the annotation pipeline tool Maker [79] which aligns mRNA, EST and protein information from the same species or cross-species to aid in gene structure determination and modifications. A consensus gene set from the above prediction algorithms was generated, using a previously described, logical, hierarchical approach [69]. In summary, the following Quality Index (QI) criteria were calculated: i) length of the 5’ UTR; ii) fraction of splice sites confirmed by an EST alignment; iii) fraction of exons that overlap an EST alignment; iv) fraction of exons that overlap EST or protein alignments; v) fraction of splice sites confirmed by a SNAP prediction; vi) fraction of exons that overlap a SNAP prediction; vii) number of exons in the mRNA; viii) length of the 3’ UTR and; ix) length of the protein sequence produced by the mRNA, and then the following decision making steps were followed: a) genes were screened for overlaps (<10% overlap was allowed); b) if QI[iii] and QI[iv] were greater than 0, or QI[iv] was greater than 0, then the gene was kept; c) the gene was BLASTed against Swissprot [80] (E < 1e-6). If there was similarity to a Swissprot entry, then the gene was kept; d) RPSBLAST was ran against Pfam [81] (E < 1e-3). If there was similarity to a Pfam entry, then the gene was kept; e) RPSBLAST was run against CDD [82] (E < 1e-3 and coverage > 40%). Genes that met both cut-offs were kept and f) if no hit was recorded, then a sequence similarity-based search was run against GenesDB from KEGG [83], and genes with at least a 55% identity and a bit score of 35 or higher were kept. Genes of interest (discussed in the paper) underwent manual evaluation and improvements. Gene product naming was determined by BER (http://ber.sourceforge.net). Functional domains and Gene Ontology (GO) terms were
assigned using InterProScan (v. 4.8) [84]. Genes were mapped to KEGG Orthology (KO) groups using wu-blastp (E < 1e-5). Proteins with signal peptides and transmembrane topology were identified using the Phobius web server [85], and non-classical secretion was predicted using SecretomeP 1.0 [86]. CEGMA (v.2.4) [18] was used to assess the completeness of the genome without excluding partial matches. MEGA6.06 [87] was used to estimate maximum likelihood phylogenies (LG+G model).

RNA sequencing and analysis

The same RNA samples were used to generate both Roche/454 and Illumina cDNA libraries. Both data types were used for genome annotation and the Illumina reads were used for differential expression analysis. Non-normalized oligo dT libraries for Roche/454 were generated as previously described [88]. The Roche/454 library was sequenced using a Genome Sequencer Titanium FLX (Roche Diagnostics) and the ‘sffinfo’ program was used to extract information from the SFF files. Adapter sequences were trimmed from the sequenced reads using the ‘seq-clean’ software and host and bacterial contamination was removed using Newbler’s ‘gsmapper’. For the Illumina RNA-seq library construction, total RNA was treated with Ambion Turbo DNase (Ambion/Applied Biosystems, Austin, TX) and 1μg of the DNase treated total RNA was poly(A) selected using the MicroPoly(A) Purist Kit according to the manufacturer’s recommendations (Ambion/Applied Biosystems). One ng of the mRNA isolated was used as the template for cDNA library construction using the Ovation RNA-seq (v.2) kit according to the manufacturer’s recommendations (NuGEN Technologies, San Carlos, CA). Non-normalized cDNA was used to construct multiplexed Illumina paired-end small fragment libraries as previously described [69], according to the manufacturer’s recommendations (Illumina, San Diego, CA) with the following exceptions. In summary, 500 ng of cDNA was sheared using a Covaris S220 DNA Sonicator (Covaris, Woburn, MA) to a size range of 200-400bp. Four PCR reactions were amplified to enrich for adaptor ligated fragments and index the libraries. The final size selection of the library was achieved by an AMPure paramagnetic bead (Agencourt, Beckman Coulter Genomics, Beverly, MA) cleanup, targeting 300-500bp. To produce cluster counts appropriate for the Illumina sequencing, the concentration of the library was determined by qPCR according to the manufacturer’s protocol (Kapa Biosystems, Woburn, MA). The Illumina HiSeq 2000 platform was used for generation of sequences of 100bp from samples of pooled individuals of 300–500 RS₃ or Sₖbred adult worms (S16 Table). RNA-seq reads were aligned to the Sₖbred reference assembly of T. circumcincta using STAR aligner (v.2.3.0) [89] with default parameters, following the 2-pass method [90]. Transcript abundance levels were expressed in FPKM (fragments per kilobase of exon per million fragments mapped). To infer rankings of differentially expressed genes according to their effect size, GFOLD (generalized fold change) algorithm was used with default parameters [91]. Although GFOLD provides a more consistent and biologically meaningful approach to ranking differentially expressed genes than other methods for single replicate RNA-seq experiments, an analysis of variation between replicate samples is necessary to draw sound conclusions, especially on the individual gene level. The differential expression patterns observed must be considered as only preliminary because of the study design limitations and confounding factors that called for cautious interpretation of the results. For example, the RS₃ worms were collected for RNA extraction at 9 days post drug treatment, resulting in a temporal separation between the stage of drug effect (and selection) and the stage at which gene expression was measured. In addition, potential sex-ratio heterogeneity among samples (that consisted of 300–500 mixed-sex adult worms) while expected to be close to 50/50 was not fully controlled for, which could result in expression variation in gender associated genes.
Variant analysis and introgression mapping by whole-genome sequencing of population pools

Genomic DNA samples from pooled individuals of 300–500 RS3 or S_inbred worms were subjected to Illumina GAII, GAIIx, HiSeq 2000, HiSeq 2500 and MiSeq paired-end sequencing (S16 Table). Sequencing adapters were removed using trimmomatic (v.0.33) [92], and the resulting reads were aligned to the S_inbred reference assembly of T. circumcincta using BWA-MEM (v.0.7.12) [93]. Picard (v.1.95) was used to remove duplicate reads and local re-alignments were performed around indels using GATK (v. 3.4–46) [94]. Variants were called from the RS3 and S_inbred population pools using GATK HaplotypeCaller (v. 3.4–46) [94]. In line with the developers’ recommendations for analyzing pooled DNA samples (as opposed to diploid individuals),—sample_ploidy parameter was set to 10. In addition to requiring a minimum mapping quality score of 20 and a minimum base quality score of 20 in the reference alignment, the following set of quality filters were applied to SNP calls using GATK VariantFiltration (v. 3.4–46) [94]: DP (maximum depth) > median depth+(median absolute deviation×1.4826)×3; QD (variant confidence divided by the unfiltered depth of non-reference samples) < 2.0; FS (Phred-scaled p-value using Fisher’s Exact Test to detect strand bias in the reads) > 60.0; MQ (Root Mean Square of the mapping quality of the reads across all samples) < 40.0; MQRankSum (Mann-Whitney Rank Sum Test for mapping qualities) < -12.5; ReadPosRankSum (Mann-Whitney Rank Sum Test for the distance from the end of the read for reads with the alternative allele) < -8.0. Population allele frequency was estimated based on the relative abundance of reads supporting each allele (i.e., allelic depth), and Fisher’s exact test was used to assess the statistical significance of allele frequency differences between the populations. To delineate the introgressed loci in the RS3 strain, we conducted a genome-wide scan of fixation index (FST) using nucleotide frequencies at polymorphic sites, and identified genomic regions that were most divergent relative to the parental S_inbred genetic background. Our mapping approach was based on the expectation that, after serial backcrossing and drug screening, causal and closely linked SNPs in the RS3 strain would retain the allelic profile of the anthelmintic resistant parental R_par isolate, whereas the genomic regions not associated with drug resistance would be represented by either the S_inbred genotype or a mixture of the two parental genotypes (depending on the variation in the recombination breakpoints among individuals). A SNP site was included in the FST analysis if it was supported by at least two alternative reads, did not overlap with indels, and had a minimum depth of 25× coverage in both populations. Following methods described by Kofler et al. [95] and assuming a pool size of 300 individuals in each population, FST values were estimated per site and averaged over non-overlapping 1-kb sliding windows. In addition, the number of loci meeting the depth of coverage threshold (25×) was examined for each window, and those windows with covered fraction > 0.5 (total combined length = 325.7 Mb) were included in the analysis. Position-sorted mean FST values (for each 1-kb window) were scanned for peaks after applying a kernel smoothing algorithm with adaptive bandwidth selection using the lokern package in R [96, 97] to identify blocks of genomic regions with extended linkage disequilibrium and elevated FST, while reducing the effects of sequencing error, mapping artifacts, and base-to-base variation in coverage. Outlier windows were identified based on the empirical distribution of the smoothed FST values. A z-score of 4.5 was chosen as the cutoff threshold, guided by the z-score exhibited by the β-tubulin gene (TELCIR_01271; z-score = 4.66), a widely recognized BZ-resistance conferring locus in trichostrongylid nematodes. We reasoned that β-tubulin isotype 1 could effectively serve as a “positive control” and that loci with FST values similar to or more extreme than that for β-tubulin gene represented candidate loci worthy of further investigation. Statistical enrichment of GO terms among the genes overlapping outlier regions was assessed using a
conditional hypergeometric test implemented in GOstat [98]. Gene-wise $F_{ST}$ values per protein-coding loci were calculated as the maximum smoothed $F_{ST}$ value among 1-kb windows that overlap the gene footprint (exon + intron). Using SnpEff (v.3.5), variants were annotated on the basis of their genomic location (e.g., exon, intron, intergenic, upstream/downstream (5 kb flanking regions) or splice site donor/acceptor), and their mutational effects were predicted (e.g., missense, nonsense or silent). Segregating haplotypes were reconstructed using either a phasing approach supported by sequencing reads [99, 100], or a de novo assembly-based method [101]. DNA copy number variation (CNV) was examined using CNV-seq [102] with p-value parameter set to 0.00001. Two-dimensional allele frequency spectra for RS$^3$ and S_inbred populations were produced after alleles were subsampled without replacement to a uniform coverage of 10x. Only bi-allelic sites with a minimum coverage of 10x in both populations were included in the analysis, while SNPs with high depth of coverage (within the top 5% of the empirical distribution) were excluded. In total, 14,562,483 SNPs were used to obtain frequency spectra. Based on these variants, Tajima’s D statistic was computed to assess whether allele frequency distributions deviated from neutral expectations [103].

**F2 mapping of IVM resistance loci using single worm double-digest restriction-site associated DNA sequencing (ddRAD-seq)**

A segregating F2 mapping population of *T. circumcincta* was generated as part of the work undertaken to introgress anthelmintic resistance genes into an anthelmintic susceptible genetic background. A cross was initiated by surgically transferring ~300 anthelmintic resistant adult males ($R_{par}$) and an equivalent number of susceptible late L4 stage females ($S_{inbred}$) into the abomasum of a previously worm-free kid goat. F1 eggs collected from the host feces were then cultured and ~10,000 of the resulting infective larvae were used to infect a second worm-free kid goat to produce an F2 generation. Two experiments were carried out to investigate the segregation of resistance to each class of anthelmintic, one using kid goats as the recipients and a second using lambs, with essentially the same results. We describe the lamb experiment here (S14 Table). Two groups of worm-free lambs (n = 27 per group) were administered an oral dose of ~8,000 (Group 1) or 16,000 (Group 2) of the F2 generation infective larvae each. On Day 27 post-infection, the lambs in Group 2 were treated with IVM. Those in Group 1 remained untreated. Three days later (Day 30 post-infection) individual fecal nematode egg counts (FECs) were undertaken on all animals and, using a restricted randomization procedure based on weight and FEC, each of the infection groups was subdivided into 3 equal-sized anthelmintic treatment groups (1a-1c and 2a-2c). On Day 31 post-infection, anthelmintic treatments were administered as follows: Groups 1b and 2b received BZ while Groups 1c and 2c received LEV. Groups 1a and 2a remained untreated as controls. Anthelmintic doses were calculated on the basis of individual live-weights and each dose was administered orally with a disposable syringe. For ddRAD-seq analysis, a total of 24 adult male F2 IVM-treatment survivors from the parallel, kid goat experiment (Group 2a) and 24 drug-naïve adult male (Group 1a) F2 worms also from the kid goat experiment were individually transferred into 100 μl of DirectPCR Lysis Reagent (Mouse Tail; Viagen Biotech, Los Angeles, CA), supplemented with 3% protease K (10 mg/ml; Roche) and incubated at 55°C for 16 h followed by 90°C for 1 h to denature the protease K. A 20 μl volume from each worm lysate (~10 ng DNA) was digested with EcoRI and MspI overnight, after which sequencing adapters (P1-EcoRI-inline-barcode and P2-MspI) were ligated to the fragment termini. The reaction was purified using a 0.5X and 0.7X double size selection (modified from Lennon et al. [104]) using Agencourt AmpureXP beads (Beckman Coulter, Brea, CA), and PCR amplified to incorporate index sequences for multiplexing using KAPA HiFi Real Time master mix using the following
protocol; 98°C for 2 min, followed by 14 cycles of 98°C for 15 sec, 60°C for 30 sec and 72°C for 30 sec. PCR reactions were purified using AmpureXP beads, after which the DNA concentrations were standardized using a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA) and samples pooled at equimolar concentration. Adapter-ligated and PCR amplified fragments approximately 500-600bp in length were obtained by gel size selection and purification. The ddRADseq library, supplemented with a 10% PhiX spike-in control, was sequenced using an Illumina MiSeq (reagent kit v3), resulting in 150bp single-end sequencing reads. Sequencing data were demultiplexed using process_radtags [105], and were mapped to the T. circumcincta reference assembly using BWA-MEM (v0.7.10) [93]. Local realignments were performed around indels using the GATK (v3.3–0), after which variants were called by HaplotypeCaller under default parameters. FST estimation [106] was carried out using VCFtools (v.0.1.12b) [107].

Genotyping β-tubulin alleles

An allele-specific multiplex PCR strategy based on that developed by Humbert and Elard [108] was used in a 96-reaction format to assess the presence of a Phe (TTC)/Tyr (TAC) substitution at codon 200 in the β-tubulin isotype-1 gene in individual worms from the anthelmintic susceptible (S\textsubscript{inbred}) and multiple-anthelmintic resistant (RS\textsubscript{3}) strains of T. circumcincta. Only adult male worms were used for these allele-specific reactions to avoid the possibility of DNA from sperm and/or fertilised eggs present in female worms interfering with the genotype identifications. The strategy involved the use of four primers per reaction, two of which—one forward and one reverse—were generic (allele-nonspecific), while the remaining two—again one forward and one reverse—were allele-specific (S2 Fig). Primer designs, which differed slightly from those of Humbert and Elard [108] to account for minor DNA sequence differences between the strains studied by them and those used in the present study, were as follows: generic forward ["TubGF"] 5’ CTTAGATGTTGTTGTAAGAGG 3’; generic reverse ["TubGR"] 5’ CATGTTCACGCCAACTTGC 3’; Phe-specific ["TubSASRev"] 5’ AGAGCTTCATTATCGAGAATGTTAA 3’; Tyr-specific ["TubRASFwd"] 5’ TGGTWGAAAAYACCGATGAAACRTA 3’. Note that TubRASFwd was degenerate at three nucleotide positions (5, 11 and 23) in order to accommodate the presence of SNPs in those positions in some haplotypes containing a Tyr at codon 200 (see S3 Fig). Two further primers—["TubRASH3Rev"] [5’ CTTCATTATCGAGAATGTTAA 3’] and ["TubSASH1Fwd"] [5’ CAGTTGGTTGAAATACCAGTGA 3’]—were designed to detect the presence or absence of a Glu198Leu substitution.

Real-time PCR quantification of single worm Tci-pgp-9 copy number

Adult worms from each of the above T. circumcincta strains were isolated from experimentally infected goats (approximately one week after successive treatments with all three anthelmintic families in the case of RS\textsubscript{3}), washed free of all debris in physiological saline and then transferred, in batches of either 100 male worms or ~300 mixed-sex worms, into cryovials where they were frozen in liquid nitrogen until required. Crude genomic DNA template was prepared from individual adult male T. circumcincta from the S\textsubscript{inbred} and RS\textsubscript{3} strains (96/strain) by overnight incubation in lysis solution [Viagen DirectPCR (MouseTail), 50 μl per worm with 3 mg/ml ProteinaseK] without further purification. SYBR\textsuperscript{G} Green real-time PCR assays were performed in a GeneAmp 5700 sequence detection system to compare Tci-pgp-9 gene copy number in individual male worms from the S\textsubscript{inbred} and RS\textsubscript{3} worm populations. Primers constructed for this purpose corresponded to genomic DNA sequence within the first putative inter-nucleotide binding domain of Tci-pgp-9 (i.e., Tci-pgp-9-IBDA) and were designed to amplify an equivalent 99bp product from each of the seven Tci-pgp-9-IBDA haplotypes.
identified from the $S_{inbred}$ and/or RS$^3$ strain worms. Primer sequences were as follows: forward [“IBD77RTGF”] 5’ CGHTATGGACGTGAAAAAGTCAcAGA 3’ and reverse [“IBD77RTGR”] 5’ CCAACTCAGGTCRGGGAAYGACTG 3’. Although designs of both these primers were based on well conserved Tci-pgp-9 IBDA haplotypes (see S6 Fig) it was necessary to incorporate some degeneracy to ensure perfect matches in all cases. To take variation in the concentration of genomic DNA between single worm DNA preparations, Tci-pgp-9 copy number was calculated in reference to a single copy gene. The single copy reference in this case was $T$. circumcincta β-tubulin isotype-1 using primers forward [“TUBRTGF2”] 5’ GGGCTTCCAACTGACGGTATTCTTTG 3’ and reverse [“TUBRTGR2’] 5’ GGGCTTCCAACTGACGCATTCTTTG 3’ which amplified a 122bp product from an exon in the central region of the gene. The $Tm$ for both primers was within the same range as those of IBD77RTGF and IBD77RTGR. All reactions were performed in duplicate in 96-well optical reaction plates (Applied Biosystems) using 25 μl reaction volumes which contained SYBR® Green PCR mastermix (Applied Biosystems), 0.2 μM of each gene-specific primer and 1 μl of 10-fold diluted crude genomic template. For both the target and reference genes “no-template controls” were included on each plate. Amplification conditions for the above reactions were as follows: initial incubation at 50˚C for 2 min, followed by 95˚C for 10 min to denature the template, followed by 40 cycles of 95˚C for 15 sec and 60˚C for 1 min. Following the reactions a melting curve and cycle threshold ($C_T$) value were generated for each sample. The $C_T$ value indicates the fractional cycle number at which the amount of amplified DNA reaches a fixed threshold. Mean $C_T$ values of duplicate samples were used in subsequent quantification analyses. No product was amplified in the “no-template control” reactions. As indicated above the amount of target measured in each case was standardised in relation to an endogenous reference gene to account for any between-worm variation in the total amount of gDNA template available. This was done by calculating Δ$C_T$ values for each sample—the Δ$C_T$ value indicates the difference in cycle number required to reach the fixed threshold for the target and reference genes. Two-sample t-tests were used to compare Δ$C_T$ values for worms from the $S_{inbred}$ and RS$^3$ populations, as well as specific genotype groups within the RS$^3$ population.

Genotyping of individual $S_{inbred}$ and RS$^3$ worms using allele-specific PCR

Tci-pgp-9-IBDA haplotypes were identified based on PCR clones (n = 66) amplified from gDNA preparations (see S6 Fig). “Allele-specific” primers were designed to differentiate between each of the haplotypes using a nested PCR strategy to allow genotyping of individual male worms from the $S_{inbred}$ and RS$^3$ strains of $T$. circumcincta. Primary reactions were performed using the degenerate primers based on the following amino acid sequences: VEIDKI-NIE (sense) [“IBD77GF3”] and GTQMSGGQ (antisense) [“IBD77GR2”]. Reactions were carried out in a Mastercycler thermal cycler (96 well block) using 20 μl reaction volumes containing 0.5 unit Platinum Taq polymerase (Invitrogen), 2 μl 10x Taq buffer, 2.5 mM MgCl$_2$, 200 μM each dNTP, 20 pmol of each primer and 2 μl template. A touchdown protocol was used as follows: 95˚C for 8 min to denature the template and activate the enzyme, followed by 12 cycles of 94˚C for 30 sec, 58˚C (~0.5˚C/cycle) for 30 sec and 72˚C for 1 min, followed by 28 cycles of 94˚C for 30 sec, 52˚C for 30 sec and 72˚C for 1 min, and finishing with a final elongation step of 72˚C for 7 min. A single generic (“allele-nonspecific”) forward primer—“IBD77GF5” [5’ GAGTATKTCACARGARCCNATGCT 3’]—was used for all subsequent nested allele-specific reactions. This primer was degenerate at four nucleotide positions (8, 14, 17 and 20) to accommodate the presence of SNPs at those positions in some haplotypes. The allele-specific reverse primers used in combination with IBD77GF5 to assess worm genotypes are shown in S17 Table. Allele-specific reactions were similarly performed in a Mastercycler using 20μl
reaction volumes but unlike the first round reactions they contained 1.0 mM MgCl$_2$ and 10 pmol of each primer, and amplification conditions used were more stringent than for the first round reactions, i.e., 95˚C for 8 min, followed by 35 cycles of 94˚C for 30 sec, 60–61˚C for 30 sec and 72˚C for 1 min, finishing with a final elongation step of 72˚C for 7 min. Five microliters of each reaction were run on a 2% agarose gel in the presence of ethidium bromide to assess the incidence of each of the sequence variants in each of the worm populations. Tci-pgp-9- IBDA genotype information derived from the allele-specific reactions was checked and verified in selected worms from each population by sequencing PCR fragments amplified from these worms using a nested protocol similar to that used for the allele-specific reactions.

Comparison of cDNA sequences encoding the N-terminal and C-terminal transmembrane regions of Tci-PGP-9 protein molecules from S$_{inbred}$ and RS$_3$ strain worms

Total RNA was isolated from mixed-sex batches of adult worms from each of the S$_{inbred}$ and RS$_3$ strains using TRI REAGENT LS (Molecular Research Center, Cincinnati, OH). Synthesis of first-strand complementary DNA (cDNA) was carried out using SuperScript II Reverse Transcriptase (Invitrogen) and poly(A) oligo(dT)$_{12-18}$ primer (Invitrogen) as per the manufacturer’s instructions. The resulting cDNA solution was diluted with DEPC-treated water to equate to an initial RNA concentration of 20 ng/μl before being stored at -20˚C until required for subsequent PCR. Overlapping fragments, encoding the complete transmembrane region from each half of the *T. circumcincta* PGP-9 protein molecule, were amplified from first-strand cDNA derived from S$_{inbred}$ (two separate pools) and RS$_3$ worms using degenerate primers in nested or partially nested PCRs. Primer designs were based on the deduced amino acid sequence of Tci-PGP-9, and corresponded to: N-terminal transmembrane region, fragment 1, first round reactions–DAIILVCFQ (sense) ["PGP9AF"]/ MIICGAFI (antisense) ["PGP9AR"]; nested reactions–VCFQFRYT (sense) ["PGP9A_fnest"]/ APFMIICG (antisense) ["PGP9AR_nest"]; N-terminal transmembrane region, fragment 2, first round reactions–GGFIVAFT (sense) ["PGP9BF"]/ YNPADGKI (antisense) ["PGP9BR"]; nested reactions–IVAFTYDW (sense) ["PGP9B_nest"]/ GCGKSTII (antisense) ["PGP9BR_nest"]; C-terminal transmembrane region, fragment 1, first round reactions–VTEDTGVA (sense) ["PGP9CF"]/ QAIOQMKFM (antisense) ["PGP9CR"]; nested reactions–ATAQNDP (sense) ["PGP9CF_nest"]/ PGP9CR; C-terminal transmembrane region, fragment 2, first round reactions–IALYFGW (sense) ["PGP9DF"]/ GCGKSTVI (antisense) ["PGP9DR"]; nested reactions–LYFGWQMA (sense) ["PGP9D_nest"]/ VGPSGCG (antisense) ["PGP9DR_nest"]. Approximate locations of these primer sites in relation to each other and to the expected transmembrane structures and the ATP sites of the PGP-9 protein molecule are depicted in S7 Fig. Although appropriate products were amplified in each case using the above primer combinations, it subsequently became apparent that there were errors in the design of the sense primers PGP9AF and PGP9AF_nest. These in fact should have corresponded to VPKASIGQ and IGQLFRYT respectively as indicated in S9 Fig. All PCR amplifications were performed in an MJ Research PTC-200 thermal cycler using final volumes of 20 μl containing 0.5 unit Platinum Taq polymerase (Invitrogen), 2 μl Taq buffer, 2.5 mM MgCl$_2$, 200 μM each dNTP, 20 pmol of each primer and 2 μl cDNA template. Both first round and nested reactions were undertaken using a touchdown PCR procedure as follows: 95˚C for 5 min to denature the template, followed by 12 cycles of 94˚C for 15 sec, 58˚C (–0.5˚C/cycle) for 30 sec and 72˚C for 1 min, followed by 28 cycles of 94˚C for 15 sec, 52˚C for 30 sec and 72˚C for 1 min, and finishing with a final elongation step of 72˚C for 7 min. PCR products (see S8 Fig) were ligated into a TOPO TA Cloning vector (Invitrogen) and multiple clones sequenced for each product.
Supporting information

**S1 Fig.** Reconstructed haplotypes of β-tubulin *isotype-1* in *Teladorsagia circumcincta* RS3 population.
(PDF)

**S2 Fig.** Allele-specific multiplex PCR strategy to assess the presence of a F200Y (tTc/tAc) substitution in the β-tubulin *isotype-1* gene in individual male *Teladorsagia circumcincta*.
(PDF)

**S3 Fig.** ClustalW multiple alignment of selected gDNA sequence variants from the central region of the β-tubulin *isotype-1* gene from *S* *inbred* and RS3 worms showing amino acid codon positions 167, 198 and 200, SNPs and positions of introns.
(PDF)

**S4 Fig.** Prediction of signal peptide cleavage site in Tci-LGC-54 (TELCIR_00170).
(PDF)

**S5 Fig.** Distribution of ΔC(T) values of individual male worms from the *T. circumcincta* strains RS3 and *S* inbred.
(PDF)

**S6 Fig.** Multiple-alignment of partial sequences of Tci-pgp-9-IBDA haplotypes showing introns, amino acid translation and locations of allele-specific reverse primers used in genotyping reactions.
(PDF)

**S7 Fig.** Approximate locations of the oligonucleotide primer sites used to amplify cDNA fragments encoding the N-terminal and C-terminal transmembrane regions of Tci-PGP-9 protein molecules from worms from the *S* inbred and RS3 strains.
(PDF)

**S8 Fig.** Agarose gel showing cDNA products amplified by PCR from the N-terminal transmembrane region of Tci-pgp-9.
(PDF)

**S9 Fig.** Deduced amino acid sequence of Tci-PGP-9 showing relative positions of amino acid substitutions associated with the multiple-anthelmintic resistant RS3 strain of *Teladorsagia circumcincta* relative to the anthelmintic susceptible *S* inbred counterpart.
(PDF)

**S10 Fig.** Fixation index (*F*ST) of ddRAD-seq derived SNP markers between IVM-screened and drug-naïve F2 mapping populations of *Teladorsagia circumcincta*.
(PDF)

**S1 Table.** Summary of *Teladorsagia circumcincta* genomic features and comparison to other clade V nematodes.
(PDF)

**S2 Table.** Amino acid composition (%).
(PDF)

**S3 Table.** *Teladorsagia circumcincta* repeat library characterization.
(XLSX)
S4 Table. Length of the Teladorsagia circumcincta genome and coding sequences (CDS) that are considered “callable” based on mapping coverage and quality.

(PDF)

S5 Table. Number of fixed and segregating bi-allelic SNPs in Teladorsagia circumcincta S_inbred and RS^3 strains.

(PDF)

S6 Table. Outlier regions (z-score > 4.5) identified through a genome-wide scan of fixation index (F_ST) between S_inbred and RS^3 populations of Teladorsagia circumcincta.

(XLSX)

S7 Table. Fixation index (F_ST) values, copy number variation (CNV) ratios, and RNA-seq transcript abundance ratios per gene.

(XLSX)

S8 Table. Nonsynonymous variants in F_ST-outlier candidate genes.

(XLSX)

S9 Table. Prevalence of β-tubulin genotypes in inbred anthelmintic susceptible (S_inbred) and multiple-anthelmintic resistant (RS^3) populations of Teladorsagia circumcincta.

(PDF)

S10 Table. Summary of ΔC_T values in relation to the occurrence of Tci-pgp-9-IBDA haplotypes in male worms of the RS^3 strain of Teladorsagia circumcincta.

(PDF)

S11 Table. Fixation index (F_ST) of ddRAD-seq derived SNP markers between IVM-screened and drug-naïve F2 mapping populations of Teladorsagia circumcincta.

(XLSX)

S12 Table. Genes displaying a minimum F_ST z-score of 2.5 in both the introgression and the F2 mapping experiments.

(PDF)

S13 Table. IVM selection signals in the introgression F_ST outlier genes.

(XLSX)

S14 Table. Summary of experimental design showing lamb treatment groups, T. circumcincta infecting doses, final group sizes (N), ivermectin treatment status and 2˚ anthelmintic treatments applied.

(PDF)

S15 Table. Between-infection-group comparisons of post-mortem worm burdens, following oxfendazole or levamisole treatment, in host animals infected with ivermectin-screened or unscreened R_par x S_inbred F2 generation Teladorsagia circumcincta.

(PDF)

S16 Table. Summary of sequenced gDNA and mRNA libraries.

(PDF)

S17 Table. Allele-specific reverse PCR primers used in Tci-pgp-9-IBDA genotyping reactions.

(PDF)
Acknowledgments

We gratefully acknowledge the assistance of the faculty and staff of AgResearch and the McDonnell Genome Institute at Washington University who contributed to this study.

Author Contributions

Conceptualization: SAB WNG MM.

Data curation: KHP JM.

Formal analysis: YJC SAB SRD KHP JM.

Funding acquisition: SAB SRD WNG MM.

Investigation: YJC SAB SRD.

Resources: SAB WNG MM.

Supervision: WNG MM.

Visualization: YJC SAB.

Writing – original draft: YJC SAB SRD JM WNG MM.

Writing – review & editing: YJC SAB SRD WNG MM.

References


50. Lynagh T, Lynch JW. Molecular mechanisms of Cys-loop ion channel receptor modulation by ivermec-


