Trends in Parasitology

CellPress REVIEWS

Opinion

Genome-wide Approaches to Investigate Anthelmintic Resistance

Stephen R. Doyle [1,*] and James A. Cotton[1]

The rapid evolution of anthelmintic resistance in a number of parasites of livestock and domesticated animals has occurred in response to widespread use of anthelmintics for parasite control, and threatens the success of parasite control of species that infect humans. The genetic basis of resistance to most anthelmintics remains poorly resolved. Genome-wide approaches are now accessible due to recent advances in high-throughput sequencing, and are increasingly applied to characterize traits including drug resistance. Here, we discuss why traditional candidate gene studies have largely failed to define the genetics of resistance, and why — and in what circumstances — we expect genome-wide approaches to shed new light on the modes of action and the evolution of resistance to anthelmintic compounds.

Highlights

Helminth control depends on the large-scale use of a limited range of anthelmintic drugs. For many species, this is predominantly a single class of drugs, which is likely to impose a strong selection pressure for drug-resistance alleles to evolve and spread.

Drug resistance is established in many veterinary helminths. Although the mode of action of some drugs is understood, the genetic basis of resistance is less well understood, despite extensive interest and research effort.

Genome-wide approaches are becoming more affordable, and well-annotated genome resources are available for many helminth species.

Genomic approaches are rapidly being implemented to characterize resistance, and they offer many advantages over traditional candidate-gene studies. However, careful consideration to experimental design is required to ensure that genetic differences discovered in these genome-wide analyses are correctly attributed to resistance.

Anthelmintic Resistance Is a Problem, and New Approaches Are Needed to Understand It

Control of both animal and human worm infections depends almost exclusively on anthelmintic drugs. Most livestock in industrialized countries are routinely treated, and while we are not aware of exact figures, at least hundreds of millions of sheep, goats, and cattle are given anthelmintic treatments to control or prevent infection every year. Human helminthiases are the target of extensive mass drug administration (MDA) campaigns, in many cases aimed at eliminating these diseases as public health problems in the coming decade. For example, the soil-transmitted helminths infect around 1.5 billion people worldwide [1] and are currently targeted with the world's largest drug administration program, with over 500 million children receiving mebendazole or albendazole (related benzimidazole drugs) in 2016 [2]. While other drug classes are available, most veterinary and human treatment relies on just three classes of drugs: macrocyclic lactones (such as ivermectin and moxidectin) and benzimidazoles for nematode infections, and praziquantel for schistosomiasis and other platyhelminth infections. The widespread use of anthelmintics to control helminths has come at a significant cost: since the first report of drug resistance in livestock helminths over 50 years ago, resistance has become far more widespread in most species of hosts and parasites [3], and in some veterinary parasites resistance to all major classes of anthelmintics has now been observed [4]. In the case of MDA campaigns, such large-scale treatment with single drug classes seems likely to promote the appearance of resistance, and there is increasing evidence for reduced efficacy of some anthelmintic classes in worms infecting humans [5–10].

The appearance of anthelmintic resistance has driven an increasing interest in characterizing the genetic mechanisms by which resistance evolves. Understanding this evolution should provide greater insight into novel approaches to control these parasites, opening the doors to molecular diagnostics for tracking drug resistance, the design of treatment strategies to avoid the appearance of resistance, and the development resistance-breaking derivatives of existing compounds [11]. This effort has largely relied on genotyping isolates of helminths with different

[1]Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK

*Correspondence: stephen.doyle@sanger.ac.uk (S.R. Doyle).

drug-response phenotypes at **candidate genes** (see Glossary), chosen based on a hypothesis about the mechanism of action of the drug [12]. The success of candidate gene approaches to date has, however, been limited: many candidate genes have been proposed to be associated with different drug classes, but few have been conclusively associated with resistance [12,13]. Although many parasites are treated with one or a few of a limited number of drug classes, high population genetic diversity, combined with their broad diversity of life history traits [14], makes it likely that different parasite species may not respond in the same way to the same drug class, and in turn, are likely to evolve unique mechanisms to tolerate exposure to anthelmintics.

The increasing interest in genetic characterization of anthelmintic resistance has coincided with remarkable increases in the throughput and decreases in the cost of genotyping technologies (see Box 1 for a summary of technologies used to assay genome variation), and particularly with high-throughput DNA sequencing [15]. Large-scale whole-genome sequencing is now accessible and is being rapidly adopted for the study of nonmodel organisms, including disease-causing pathogens, and the characterization of traits such as drug resistance [16–23]. Similarly, genome-wide approaches have begun to be adopted by helminth parasitologists to investigate traits including anthelmintic resistance [24–29]. These studies have coincided with a rapid increase in draft and reference genome sequences for helminths [14], as well as publicly available resources to support researchers in making best use of these data [30,31]. The increasing accessibility of genome-wide data highlight that the interpretation and significance of single candidate gene studies need to be carefully considered. At the same time, viewing genome-wide studies through the lens of single candidate gene studies also brings a new perspective on why careful interpretation of their results is warranted. In this Opinion article, we discuss some of the opportunities and considerations for the use of genome-wide approaches to investigate anthelmintic resistance.

How Informative Have Candidate Genes Been, Really?

Much of what is known about the genetics of anthelmintic resistance is based on candidate gene studies. The textbook example is β-tubulin and its association with benzimidazole anthelmintics: resistance-causing mutations were first characterized in *Aspergillus nidulans* [32,33], and then identified in *Caenorhabditis elegans* [34]. One or more of three specific nonsynonymous substitutions in β-tubulin isotype 1 have subsequently been shown to be associated with treatment failure in most nematodes targeted with benzimidazole-related anthelmintics [35,36]. The validation of β-tubulin mutations and their impact on resistance in parasitic helminths that are targeted by benzimidazoles is undeniably important, but this single example can hardly be held up as general support for *ab initio* candidate gene studies identifying resistance-associated loci in parasitic nematodes.

Candidate gene approaches have been less successful in deciphering the resistance mechanism(s) for other anthelmintics. Many reports describe variation in gene expression or allele frequencies between phenotypically susceptible and resistant populations, often concluding that selection has occurred at the examined candidate gene. For example, at least 30 loci have been proposed as potentially involved in ivermectin resistance (see Supplementary Table 2 in [27]). However, the association between candidate genes and resistance often does not replicate between different species or even in independent studies of the same species [24,26,27,37–39]. These include, for example, experimentally defined major effect loci for ivermectin resistance in *C. elegans* such as avr-14, avr-15, and glc-1 [40], yet there is little if any evidence for positive selection associated with resistance in these genes in any parasitic nematode. To date, none of the current candidate genes are used for diagnosis of resistance in

Glossary

Bulk-segregant analysis: a genetic mapping approach to identify markers that are associated with a phenotype, by comparing genetic variation between pools of samples that show extremes of the phenotype of interest.
Candidate gene: a gene proposed to be associated with a given trait. Candidate genes are typically chosen based on known or *a priori* assumption of function.
Genetic architecture: describes the underlying genome-wide distribution of genetic variation in individuals and/or populations, and the contribution of these variants to phenotypic variation.
Linkage disequilibrium (LD): describes combinations of alleles that occur more frequently than expected from the frequency of the individual alleles alone. LD can be very informative about processes maintaining genetic structure in a population.
Linkage mapping: a technique used to identify regions of the genome that segregate with a phenotype of interest in the progeny of a genetic cross between two phenotypically distinct parental isolates.
Manhattan plot: a type of scatter plot typically used to present genome-wide data. Each point on the plot represents a position (typically a variant position or window region) in the genome on the x axis, and test statistic of a comparison between two or more groups on the on the y axis.
Population structure: describes the nonrandom distribution of genetic variation between two or more subgroups of individuals.
Principal component analysis: a dimension reduction approach that can be used to convert large and/or multivariate datasets into ranked simplified components that describe variability in the data. Useful for detecting population structure and other variables that may influence the structure of genomics data.
Quantitative trait locus (QTL): a region of the genome containing one or more genes that are linked to a trait of interest in a population. QTLs are identified by the correlation of particular genotypes with a

the field. One key problem is that few studies take into account the broader genomic landscape beyond a single (or few) gene(s): as we will discuss, the focus on single candidate genes in isolation can be misleading without knowledge of the surrounding genomic context, and therefore it is difficult to conclusively discriminate between selection and confounding biases (Box 2) that may arise when comparing genetically distinct strains [13,41]. Importantly, the focus on individual candidate genes prevents the discovery of novel genes and mechanisms that might be at play. The association between specific mutations of β-tubulin isotype 1 and benzimidazoles is a case in point: although much of the literature is focused on this specific association alone, recent genome-wide analyses using *C. elegans* laboratory strains and wild isolates revealed β-tubulin isotype 1 (*ben-1*)-independent **quantitative trait loci** (QTLs) that differentiated benzimidazole analogues [28] and multiple novel *ben-1* specific alleles [42] associated with resistance, respectively. The impact of noncanonical β-tubulin mutations on benzimidazole resistance in parasitic species is unknown, but clearly deserves further attention. Importantly, a healthy dose of scepticism is warranted in the interpretation of candidate genes and their role in anthelmintic resistance, and unbiased approaches that account for the impact of selection on the broader **genetic architecture** are needed to properly understand the genetics of anthelmintic resistance.

Defining the Observed and Expected Genetic Architecture Associated with Anthelmintic Resistance

What should the genetic architecture of anthelmintic resistance look like in a genome? Heritable resistance relies on the presence of specific alleles of genes that alter how susceptible individuals, carrying those alleles, respond to drug treatment. The spread of these alleles depends on individuals that contain one or more resistant alleles having a fitness advantage upon drug exposure over those that do not. In populations composed of both susceptible and resistant individuals, the expected pattern of genetic variation throughout the genome after drug treatment depends on the frequency of the resistance-conferring alleles before treatment, the number of different alleles that can generate a resistance phenotype, and the strength of selection on those alleles. Resistance-conferring alleles are conventionally thought to arise *de novo* in a population and be selected for from a very low starting allele frequency in response to drug treatment. If resistance alleles have a large 'effect size', that is, if individuals with these alleles survive and reproduce while those without them die or do not reproduce, then drug treatment will exert a strong selective pressure, resulting in a rapid loss of susceptible alleles and an increase in frequency of the novel resistance alleles in the population as susceptible individuals are replaced by resistant ones. In this context, often termed a 'hard sweep', the genetic diversity surrounding the resistant allele would be significantly reduced, and due to the rare starting allele frequency of the resistant mutation, the genome-wide diversity would also be significantly reduced to reflect the limited genetic diversity in the few individuals that harboured the mutation (Figure 1A–C).

In contrast, resistance alleles may coexist on diverse genetic backgrounds in the population as a result of older or pre-existing variation (i.e., prior to drug pressure), or via the recurrence of resistance mutations in the same population. In this case, a 'soft sweep' occurs, associated with a reduction in the genetic diversity around the causative locus after drug treatment, but unlike a hard sweep, the background genetic variation is likely to be maintained to some degree (Figure 1A,D,E). A similar pattern will occur if there is selection acting on multiple, genetically distinct mutations that each confer resistance. Where multiple genes are driving resistance directly, each allele may have an additive effect towards resistance (e.g., the combined effect of two alleles is equivalent to the sum of effects from individual alleles), or show synergistic interactions, whereby the effect of combinations of alleles is greater than the sum of their individual effects. Considering the large population sizes and extensive genetic diversity of

phenotype. As QTLs are often associated with continuous traits, the number and shape of QTLs inform the genetic architecture of the trait.
Recombination: the process by which DNA is rearranged between homologous chromosomes during meiosis. Recombination can produce novel combinations of alleles in the gamete relative to either parental chromosome, the result of which is to generate phenotypic variation in the offspring.
Statistical power: describes the probability that a statistical test will correctly reject a null hypothesis when the alternate hypothesis is true. Power ranges from 0 to 1: the higher the value, the less likely type II or false-negative errors will be made, whereas power values closer to 0 are associated with higher type I or false-positive errors.
Type I and Type II errors: describe when the null hypothesis is rejected when it is true (a type I error is a 'false positive', for example, an SNP that has achieved genome-wide significance, but is not associated with the trait), and accepting the null hypothesis when it should in fact be rejected (a type II error is a 'false negative', for example, an SNP that has not reached genome-wide significance, but is associated with the trait), respectively.

Box 1. Approaches to Assaying Genome-wide Genetic Diversity

The resolution to detect genetic changes associated with traits such as anthelmintic resistance is largely dependent on the assay or technology used. From sequencing whole genomes to targeted sequencing of individual genes, we outline some of the benefits and limitations of a number of approaches that could be used to assay variation in Table I.

- Whole-genome sequencing of individuals is increasingly being employed for genetic association analyses; although this approach provides maximal information at single-nucleotide resolution, allowing correlation between genotype and phenotype, this approach may require significant economic and analytical/computational resources, particularly as the genome and sample size increase.
- Pooled whole-genome sequencing provides a compromise – DNA from multiple individuals is pooled and sequenced, and so while individual genotype information is lost, allele frequency estimates of the pool can be determined and compared between pools. Note that significant technical variation can occur between pools, and therefore, achieving sufficient sequencing depth per pool to sample the allelic diversity and replication of pools is important [70].
- RNAseq and exome-based approaches aim to sample a proportion of the coding regions of the genome, with the former sampling actively transcribed mRNA at the time of sample collection, whereas the latter aims to sample genomic DNA by using 'baits' to capture targeted regions of the genome. Capture-based approaches are not limited to the exome, and can be designed to target small to large regions of the genome [Helminth exomes: range = 7 (*Parascaris equorum*) to 96 Mbp (*Macrostomum lignano*), median = 19 Mbp]; however, such an approach requires initial investment in the capture design and bait synthesis.
- Restriction site-associated DNA sequencing (RADseq) is increasingly being used to sample genome-wide variation that occurs near restriction sites throughout the genome in a number of nonmodel organisms. However, as the distribution of restriction sites is not random, and therefore not evenly distributed throughout the genome, this approach may not detect association if a comparable restriction fragment is not within a linked region to a causative allele. *In silico* prediction of restriction sites to estimate coverage, particularly when a contiguous genome sequence is available, is recommended.
- The use of SNP genotyping platforms, such as Sequenom, microsatellites, or sequencing amplicons from single or small numbers of genes, are best utilized to fine-map regions, and are not particularly suitable for discovery. Each assay requires knowledge of genomic coordinates in the experimental design, and therefore are generally impractical to apply on broad region(s) of the genome while maintaining sensitivity to detect associations. In combination with a genome-wide approach to define the broader context of variation, these approaches allow confirmatory experiments and a wider-scale investigation of geographical distribution of a specific signature of selection in the genome.

Table I. Consideration of Approaches to Assay Genetic Variation Associated with Resistance[a]

Genotyping approach	Whole-genome sequencing	Pooled WGS	Exome/capture arrays	RNAseq	RADseq	Sequenom	RFLP	Amplicons genotyping/ RT-qPCR	Single genes
Assay approach summary	DNA > WGS of individuals	DNA > WGS of pools of individuals	DNA > probe-based DNA sequence capture > HT sequencing	RNA > cDNA > HT sequencing	DNA > restriction digest > HT sequencing	DNA > primer extension PCR > MALDI-TOF Mass spectroscopy	PCR > Electrophoresis	DNA/ RNA > PCR > HT or Sanger sequencing OR RNA > RT-qPCR	DNA > PCR > Sanger sequencing
Throughput	Low/moderate	Low pools, high samples	Low/moderate	Low/moderate	Low > High	Low/moderate	Low/moderate	Low > High	Low > High
Cost per individual	High	Low	Moderate	Moderate	Low	Moderate	Low	Low	Low
Number of variants	High	High	Moderate	Moderate	Moderate	Low/moderate	Low	Low	Low
Cost per variant	Low	Low	Low	Low/moderate	Low/moderate	Moderate	High	High	High
Resolution for assaying genetic variation linked to resistance	High	Moderate/high	High	Moderate	Moderate	Low/moderate	Low	Low	Low
Helminth references	[29,58]	[24–27]	[71]	[72,73]	[25,74,75]	[24,26]	[11,66]	[76–79]	Many

[a]WGS, whole-genome sequencing; HT, high throughput; RADseq, restriction site-associated DNA sequencing; RFLP, restriction fragment length polymorphism.

Box 2. Challenges with Helminth Genomes

The transition from assaying single genes to whole-genome analyses for helminths does present some biological and technical challenges.

Genome Size and Gene Density Variation

Helminths typically have genome sizes ranging from 43 Mbp (*Parastrongyloides trichosuri*) to 1.2 Gbp (*Spirometra erinaceieuropaei*) (median = 122 Mb), and are predicted to contain anywhere between 8140 (*Onchocerca flexuosa*) and 101 269 (*Meloidogyne arenaria*) coding sequences (median = 17 498). While helminth genomes are considerably smaller than mammalian genomes, the gene density is higher (7 to 507 genes per Mbp; median = 142; compared with human of 12 to 15 genes per Mbp), which will increase the number of genes linked to any given association signal.

Extensive Genetic Diversity

Helminth populations can be genetically diverse. The presence of extensive genetic diversity increases the likelihood of identifying regions of the genome that differ between isolates, for example, a genome-wide scan of differentiation between resistant and sensitive isolates demonstrates broad differentiation in *Haemonchus contortus* [27], and focus on an individual locus could have been easily mistaken for positive selection without taking into account the genome-wide context. Although increased genetic diversity can potentially provide better resolution for QTL analyses due to a higher marker density, such diversity does make it harder to identify causal variants.

Choosing a Reference Genome

The use of a reference genome sequence is a key component to most genome-wide studies. However, such analyses ultimately can only study what is known in the reference, and sequences present in the sample that are missing from the reference will not be assayed. Significant genetic diversity exists within and between helminth populations, and while single nucleotide substitutions and insertion/deletions are generally reliably assayed, more complex genomic rearrangements and copy-number variants that are likely to be present are typically much harder to characterize. Such variation may, in fact, play a role in resistance; for example, a copy number variation (CNV) expansion of *pgp-9* is suspected to be involved in ivermectin resistance in *Teladorsagia circumcincta* [25]. A single reference sequence is unlikely to be able to account for large-scale rearrangements, and therefore, the *de novo* assembly of multiple genetically distinct reference genomes may be required to correctly characterize such variation.

Genome Contiguity and Completeness

From a technical perspective, the contiguity of the genome assembly will play a major role in detecting associations. Although there are now a number of helminth species with chromosomal-scale genome assemblies (i.e., *Brugia malayi, Haemonchus contortus, Onchocerca volvulus, Schistosoma mansoni,* and *Trichuris muris*; www.parasite.wormbase.org), the genome assemblies for most helminths remain highly fragmented. Poorly assembled genomes almost certainly contain artefacts that do not reflect the actual genome of the organism, for example regions of the genome inaccessible to sequencing, multiple haplotypes, collapsed paralogs, contaminants, and poorly resolved repetitive regions. All of these artefacts will impact the utility of the genome for genome-wide studies. The increasing use of long-read sequencing technologies to assemble helminth genomes is an important step towards resolving some of these issues; however, even with long-read sequencing technologies, high-quality chromosomal-level assemblies will remain challenging to produce (and need investment) for the foreseeable future.

many helminths targeted by anthelmintics, there are likely to be many alleles of some effect that influence the response to anthelmintics segregating in populations. These include alleles that confer complete insensitivity to the drug, but also alleles that influence the pharmacokinetics and therefore exposure of the individual to the drug. The extent of these interactions is dependent on the frequency of each resistance allele which, if different between populations, will produce distinct signatures of selection that further complicates the detection and interpretation of resistance.

In this context, the genetic architecture of anthelmintic resistance can be complex, and is likely to be shaped by a combination of both hard and soft selection. Interpreting these signatures of selection at a candidate gene can be challenging without first understanding and accounting for the surrounding genomic variation. Genome-wide approaches seem to be the most sensible way to do this; however, genome-wide data can be misled by biological and technical variation or poor experimental design, leading to confounding and/or lack of **statistical power** to confidently detect an association. Some of these issues are discussed below.

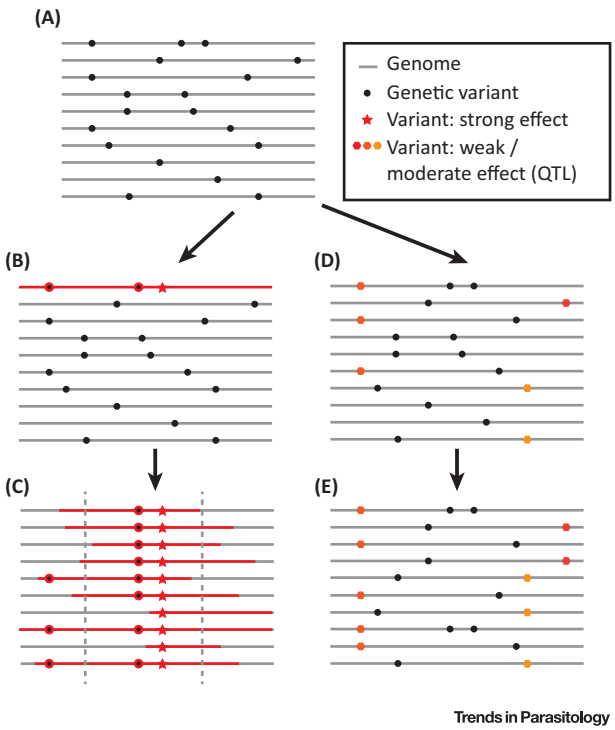

Figure 1. Defining the Expected Genetic Architecture of Resistance. The distribution of genetic variation across the genome after selection is dependent on the number and frequency of causative resistance variants present in a population (A). In the classic example of selection for drug resistance, a rare allele of large effect will appear in a population (B), and upon selection, will rapidly increase in frequency and may become fixed in the population. Termed a 'hard sweep', any variant linked to the resistant allele will also increase in frequency, reducing the genome-wide variation of the population (C). In contrast, large effect resistance-associated alleles present on multiple genetic backgrounds, multiple alleles of smaller effect, or perhaps alleles associated with tolerance of drug exposure (D), will each increase in frequency after drug exposure. (E) A 'soft sweep', unlike a hard sweep, will retain a larger proportion of the pretreatment genetic variation maintained in the population. Abbreviation: QTL, quantitative trait locus. Figure adapted from Doyle *et al.* [24].

Defining Robust Associations between Genetic Variation and Anthelmintic Resistance

Characterizing the genetic basis of resistance involves the statistical differentiation of genotype and/or allele frequencies between groups of resistant and susceptible individuals, or proxies such as treated and untreated worm populations. Whether assaying a single gene or whole genomes, often simple statistical approaches such as chi-square tests or regression approaches are used, although more complex approaches have been employed; the exact statistical approach differs in whether the phenotype is quantitative (e.g., EC50 values) or qualitative (e.g., survival vs death), and whether individual worms or pools of worms have been assayed (e.g., **Bulk-segregant analyses**). However, moving from detecting an association at a single gene to testing all variants in the genome requires more robust statistical testing [43]. The most obvious requirement is to determine a genome-wide threshold for which a statistical test is deemed statistically significant: such a threshold needs to account for the increase in the number of statistical tests used, and for the likelihood that a proportion of tests will appear to be significant by chance alone. For example, a *P* value threshold of 0.05 for independent tests of 1 million variants is expected to identify 20 000 variants, reaching that significance value even if the null hypothesis is true in every case, representing many **type I 'false-positive' errors**. In human association studies, the threshold *P* value for an individual test is typically taken as $P = 5 \times 10^{-8}$. For other organisms, the data itself, and the number of tests used, should influence the choice of threshold: formal multiple-testing correction can be calculated using a variety of approaches including, but not limited to, using a Bonferroni correction or similar to control the chance of a single type 1 error among all the tests (family-wise error rate; note that this approach is typically conservative and may lead to **type II 'false-negative' errors**), or instead controlling the proportion of tests with significant *P* values that are false rejections of the null hypothesis (the false discovery rate), which may be more realistic when many hypothesis

tests are performed [44]. These thresholds are typically calculated using the total number of variants in the genome, and each variant is considered independently; however, as **recombination**, and therefore **linkage disequilibrium**, are not evenly distributed throughout the genome (Box 3), these thresholds can also be calculated from linked haplotype blocks of multiple variants, reducing the total number of statistical tests that need to be accounted for. Statistical approaches for sequence data from pooled individuals are less well defined, and often not based on an explicit null hypothesis but instead rely on choosing a more-or-less arbitrary cut-off based on either the distribution of the test statistic (e.g., *P* value, F_{ST}) to define the genome-wide significance, that is, a fixed quantile or standard deviation from the genome-wide mean, or a fixed number of SNPs, that is, the top 100 SNPs.

Box 3. Genome-wide Scans for Diagnostic Markers

Considering the challenges of identifying and subsequently validating causative mutations for complex traits, an initial step might be to identify alleles that can act as a marker for the trait. Such a marker might be used as a diagnostic tool to predict the appearance or frequency of a trait, and inform, for example, treatment practices. A diagnostic marker need not be the causative variant, but does need to sufficiently correlate with the trait, that is, the frequency of the marker is high in individuals with the trait, which may occur via linkage between the diagnostic marker and the causative variant due to it being physically located nearby to the causative variant. One example is the microsatellite ms8a20 and its association with ivermectin resistance in *Haemonchus contortus*: the association was originally identified in a genetic cross [66], and has been subsequently validated in independent field studies (John Gilleard and Andrew Rezansoff, personal communication).

The extent of linkage surrounding a causative mutation is influenced by mechanisms that break down linkage, that is, recombination, and forces that maintain linkage disequilibrium, which describes the nonrandom association of alleles. Because recombination typically occurs little more than once per chromosome per meiosis during the production of gametes, the frequency of recombination events between two loci is largely dependent on the distance between them, that is, markers far apart will have a higher chance of a recombination event disconnecting them than markers that are close to each other. The frequency of recombination events, and in turn the pattern of linkage disequilibrium, are however not randomly distributed throughout the genome, and therefore, the diagnostic potential of a marker not only relies on the distance but also on the precise location of the marker and causative variant in the genome.

The genetic diversity of a population is ultimately shaped by processes that include recombination and linkage disequilibrium. This is relevant for the broader diagnostic utility of a marker beyond the population in which it was identified, and is particularly true for many studies that characterize variation in candidate genes between isolates that differ not only in their resistance phenotype but also spatial or temporal sampling; if the linkage structure is sufficiently different between populations (Figure IA), a diagnostic marker may vary in its linkage with the trait, to the degree in which it becomes unlinked due to sufficient recombination (Figure IB). It is therefore important that proposed diagnostic variants should be validated in independent populations; additional validation will either (i) increase support for a marker, and provide an estimate on the variance in prediction, or (ii) decrease support, either due to technical confounding of the original prediction, that is, it was not a good marker to begin with, or that there may be additional genetic loci involved that differ between populations.

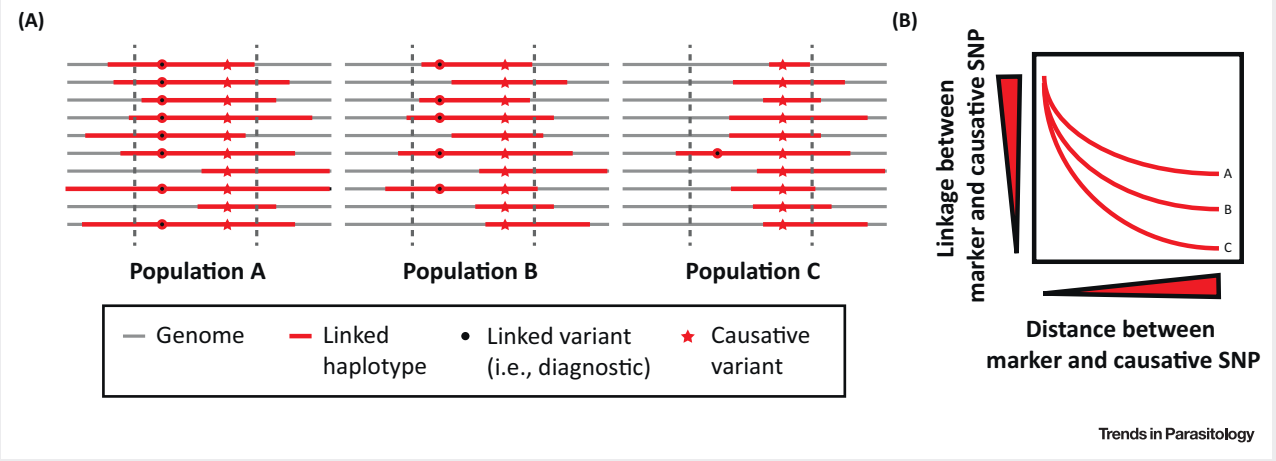

Figure I. Impact of Linkage and Recombination on the Association between a Causative and Putative Diagnostic Marker of Resistance. (A) Three hypothetical populations are considered, each with a different amount of structure, from 'high' (Population A), 'moderate' (Population B), to 'low' (Population C) genetic linkage where haplotypes are indicated by the length of the red line. (B) Although the distance between causative and diagnostic SNPs remains fixed, the amount of linkage in the population influences the correlation of association between the two positions, and therefore influences the diagnostic potential of a marker, particularly if it is applied to independent populations from which it was originally defined.

Visualizing the patterns of linked variation throughout the genome can provide additional context beyond a ranked list of variants that achieve a genome-wide level of significance. Identifying regions of the genome where a test statistic (i.e., *P* value, F_{ST}) from multiple adjacent variants behaves similarly in a way consistent with a sweep acting on and around a causative gene, and all achieve genome-wide level of significance, provides more convincing evidence of selection than a sporadic distribution of significant test statistics throughout the genome. The ability to visualize signals of linked variation, typically performed using a **Manhattan plot**, is dependent on the density of markers sampled and the amount of linkage disequilibrium between markers in the particular region of the genome (see Box 3 for a discussion of linkage and its importance for developing genetic markers of resistance), and the contiguity of the reference sequence used in the analysis. Putative signals of genetic differentiation can be misleading in genome-wide approaches if the reference genome used is very fragmented – technical artefacts aside (Box 2), many small apparent QTLs identified in a fragmented assembly could in fact be from a single, larger region that is undergoing a single selective sweep [27]. Understanding the expected distribution of genetic variation in response to selection can inform the interpretation of the observed statistical data in the context of its distribution throughout the genome, and in turn, help to distinguish false from true signals of selection.

The statistical power to detect an association largely relies on the size of the effect and number of samples used. In human genetics, association experiments increasingly involve sample sizes of tens to hundreds of thousands of individuals, a necessary requirement to detect a signal among many variants with small effects in heterogeneous sample sets [45]. The effect size of drug resistance is expected to be large, and therefore, sample sizes may not need to reach those required for the study of many human traits. However, loci with small effect sizes will be more difficult to characterize, and larger samples sizes may be required. As sample sizes increase, the phenotypic and genotypic variation in the sample cohort is also likely to increase. Precise phenotyping of individual helminths is necessary, but can be surprisingly difficult; many measures of drug efficacy are performed on pools of individuals and on life stages that are not necessarily the direct targets of drug selection, for example, egg-reduction tests in gastrointestinal nematodes [46], or the microfilariae repopulation rate for filarial nematodes [47]. While it is possible to perform *in vitro* drug selection assays, such as those based on egg hatch [48], larval migration [49], larval development [50], and larval motility [51], to examine resistance of some helminth species, they too are largely focused on the average response of multiple individuals, and it is not yet clear if the phenotypic response, that is, motility, is a sufficient proxy for survival in the presence of a drug. Both phenotypic and genotypic variation is partitioned within and between biological samples, and therefore it is important that individuals sampled within each comparison group are carefully matched, for example, by developmental stage, sex, and genetic background to maximize statistical power to detect a genetic association. The source of some of these sampling biases, and ways to correct for them, is discussed below.

Associations Should Be Interpreted against the Genetic Background of a Population

A common analytical approach to assessing anthelmintic resistance genetics has been to compare the phenotype and genotype of susceptible and resistant strains of parasites. One issue is that this kind of experimental design is overly simplistic – in some cases, there is an assumption that strains that happen to differ in drug resistance phenotype differ only in drug resistance phenotype, so that all or much of the genetic variation between them is likely to be involved in resistance. Helminths are genetically diverse, and depending on the life history traits of the particular parasite, this genetic diversity may be structured at multiple hierarchical levels (Figure 2A) [41]. As such, in the presence of high genetic diversity and unknown genetic

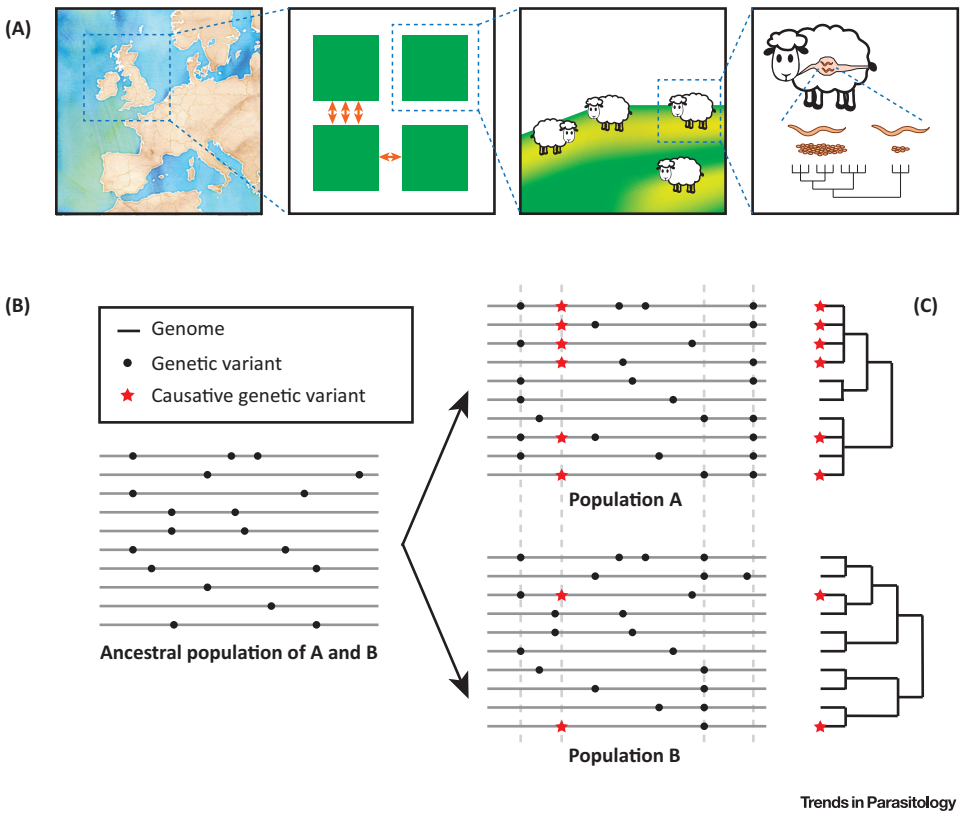

Figure 2. The Impact of Genetic Structure on Detecting Associations. (A) Genetic population structure can arise and be maintained across a number of different hierarchical levels. Between countries: variation reflects distinct evolutionary histories, likely with no or low genetic exchange. Within a country: host movement, parasite management practices, and environmental influences may differentially promote, or limit, parasite spread between regions. Within a region: familial structures, behavioural differences, and differential susceptibility to parasites (i.e., between different host species, or host genetic variation within a species) between hosts may influence the distribution of parasites. Within a host: variation in development rate and reproductive output of parasites may influence the proportion of progeny contributing to the subsequent generation of reproductively viable parasites. Comparisons of genetic variation within and between levels of this hierarchy will influence the detection of an association, and the confounding effects of such genetic structure on this association. (B) Genetic comparison of two genetically and phenotypically defined populations, A and B, that originated from a single ancestral population will share a proportion of their genetic variants. However, background genetic variation will vary, both linked to the resistance alleles and via stochastic processes such as genetic drift. In this case, many noncausal variants will be statistically associated with resistance even without genetic linkage to a causative resistance mutation. (C) Family structure and cryptic relatedness between samples has the potential to confound phenotype frequency estimates, and in turn, impact the detection of true genetic associations. Association testing between these populations without correcting for, or removing, close relatives, would detect many variants of high frequency common between relatives, which would be largely indistinguishable to any variants linked to resistance.

structure, both candidate gene and genome-wide approaches alike can be confounded by spurious genetic signals that simply reflect **population structure** rather than association with resistance (Figure 2B). In general, understanding the genetic structure of the populations that isolates come from is key to interpreting the genetic variation between them.

Fortunately, for genome-wide (or multilocus) data, the genetic data itself gives us a robust picture of the underlying genetic structure, as most loci will not be under strong selection or associated with any single phenotype of interest. Methods designed to assign individuals to discrete population clusters using genetic variation from each individual [52–54] can discover

underlying broad-scale genetic structure. Statistical approaches to correct for this structure can subsequently be used, for example, by including the position of each sample on the first few axes of a **principal component analysis** as covariates [53], or modelling expected genetic covariance due to relatedness or population structure in a number of mixed-effect model association methods [55,56]. More subtle levels of genetic relatedness such as family structure and cryptic relatedness between samples can also confound association testing (Figure 2C), as some genetic variants that are rare in a population may be common between closely related individuals. Approaches to detect such relationships have been used to reconstruct pedigrees [57] and complex familial relationships such as polyandry [58] among helminths; in the context of association analyses, these methods can be used to robustly identify relatives, which should be removed from data to avoid pseudoreplication of similar genotypes.

Without prior knowledge of the existing genetic structure for a given species, a better approach is to minimize the presence of population structure at the experimental design stage. This could be achieved by ensuring that susceptible and resistant isolates are carefully matched (just as cases and controls need to be in other association studies, as reviewed in [59]). Alternatively, a population can be built artificially for mapping by using a genetic cross between genetically distinct isolates. In this approach, the population genetic differences between parents become irrelevant in the F1 progeny, which will all be identically heterozygous for all fixed differences between the parents. By characterizing genetic markers that cosegregate with the trait of interest in the subsequent generation(s), it is possible to map trait-associated genes to discrete regions of the genome. The resolution of this approach, termed **linkage mapping**, is dependent on the amount of recombination present to break down linkage between markers from the original parental haplotypes: this can be influenced by the number of offspring analysed (or at least passaged during the cross), and/or the number of generations since the original cross. Genetic linkage mapping is of course limited to genetically tractable species for which a life cycle can be maintained. Routinely performed with *C. elegans* [28,60–62], genetic crosses and mapping experiments have been demonstrated in a number of helminths, perhaps most successfully using the human-infective parasite *Schistosoma mansoni* maintained in mice [63,64] (recently reviewed in Anderson *et al*. [65]), and increasingly in veterinary helminths including the livestock parasites *Haemonchus contortus* [27,58,66–68], *Teladorsagia circumcincta* [25], and *Fasciola hepatica* [69]. Although these examples represent a small proportion of the number of species targeted by anthelmintics, together these include species that are resistant to all of the major classes of anthelmintics, and therefore seem most relevant to confirm the role of candidate genes or identify novel resistance-associated mutations that can subsequently be explored in less genetically tractable species.

Concluding Remarks

Genome-wide approaches can provide great insight into the distribution of genetic variation throughout the genome, and how selection on genes associated with traits such as anthelmintic resistance shapes this variation. We have argued that genome-wide approaches will greatly improve the sensitivity and specificity of experiments searching for the genetic basis of anthelmintic resistance, and help to unravel the contribution of known and novel genes towards drug failure (see Outstanding Questions). The discovery, validation, and interpretation of these data is dependent on, and greatly benefits from, ongoing development of publicly available parasitological, genetic, and bioinformatic resources and databases. Working with these organisms will remain challenging, both practically in the need to collect phenotypically defined parasites from natural or experimental sources, and analytically, thanks to the high genetic diversity and unusual population genetics caused by the life cycle and reproductive biology of these organisms. However, genome-wide approaches will be successful in the context of well-

Outstanding Questions

Is it possible to clearly delineate the genetics of resistance from tolerance, and how important are both in clinically relevant drug failure?

What are the effect sizes of resistant alleles? If multiple resistance alleles are segregating in a population, can we accurately assign effect size to prioritize candidate markers of resistance?

What proportion of genetic variation is 'missing' due to the use of a single reference genome in genetically diverse species?

Phenotypic characterization is clearly important but often difficult to assess. What technological or analytical approaches are required to improve and/or standardize the characterization of resistance (and other helminth) phenotypes?

In the absence of causative resistance alleles, genetic markers that infer resistance are desired. What minimal evidence is required, and what degree of confidence is needed, to validate a marker as predictive of resistance?

designed and sufficiently powered studies of matched (or experimentally produced) cases and controls, identified by good-quality phenotypic data. In this context, genome-wide approaches promise to refine our understanding of the genetics and evolution of anthelmintic resistance well beyond what is possible with the current long list of proposed candidates available today.

Acknowledgments

The authors research is supported by Wellcome Trust (grants 098051 and, 206194) and by the Biotechnology and Biological Sciences Research Council (BB/M003949/1). We thank Roz Laing, Guillaume Sallé, John Gilleard, Neil Sargison, and members of the Parasite Genomics group at Wellcome Sanger Institute for helpful discussion and constructive comments.

References

1. Pullan, R.L. *et al.* (2014) Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites Vectors* 7, 37
2. Garba Djirmay, A. and Montresor, A. (2017) *Schistosomiasis and Soil-Transmitted Helminthiases: Number of People Treated in 2016,* WHO/Department of Control of Neglected Tropical Diseases
3. Sangster, N.C. *et al.* (2018) Ten events that defined anthelmintic resistance research. *Trends Parasitol.* 34, 553–563
4. Kaplan, R.M. and Vidyashankar, A.N. (2012) An inconvenient truth: global worming and anthelmintic resistance. *Vet. Parasitol.* 186, 70–78
5. De Clercq, D. *et al.* (1997) Failure of mebendazole in treatment of human hookworm infections in the southern region of Mali. *Am. J. Trop. Med. Hyg.* 57, 25–30
6. Ismail, M. *et al.* (1999) Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *Am. J. Trop. Med. Hyg.* 60, 932–935
7. Melman, S.D. *et al.* (2009) Reduced susceptibility to praziquantel among naturally occurring Kenyan isolates of *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* 3, e504
8. Crellen, T. *et al.* (2016) Reduced efficacy of praziquantel against *Schistosoma mansoni* is associated with multiple rounds of mass drug administration. *Clin. Infect. Dis.* 63, 1151–1159
9. Osei-Atweneboana, M.Y. *et al.* (2011) Phenotypic evidence of emerging ivermectin resistance in *Onchocerca volvulus*. *PLoS Negl. Trop. Dis.* 5, e998
10. Nana-Djeunga, H.C. *et al.* (2014) Reproductive status of *Onchocerca volvulus* after ivermectin treatment in an ivermectin-naïve and a frequently treated population from Cameroon. *PLoS Negl. Trop. Dis.* 8, e2824
11. Valentim, C.L.L. *et al.* (2013) Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. *Science* 342, 1385–1389
12. Gilleard, J.S. (2006) Understanding anthelmintic resistance: the need for genomics and genetics. *Int. J. Parasitol.* 36, 1227–1239
13. Gilleard, J.S. and Beech, R.N. (2007) Population genetics of anthelmintic resistance in parasitic nematodes. *Parasitology* 134, 1133–1147
14. International Helminth Genomes Consortium *et al.* (2018) Comparative Genomics of the Major Parasitic Worms. *Nat. Genet.* Published online November 5, 2018. http://dx.doi.org/10.1038/s41588-018-0262-1
15. Wetterstrand, K.A. (2018) DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). www.genome.gov/27541954/dna-sequencing-costs-data/ (accessed: October 9, 2018)
16. Chang, H.-H. *et al.* (2013) Malaria life cycle intensifies both natural selection and random genetic drift. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20129–20134
17. Cheeseman, I.H. *et al.* (2015) Pooled sequencing and rare variant association tests for identifying the determinants of emerging drug resistance in malaria parasites. *Mol. Biol. Evol.* 32, 1080–1090
18. Cheeseman, I.H. *et al.* (2012) A major genome region underlying artemisinin resistance in malaria. *Science* 336, 79–82
19. Park, D.J. *et al.* (2012) Sequence-based association and selection scans identify drug resistance loci in the *Plasmodium falciparum* malaria parasite. *Proc. Natl. Acad. Sci. U. S. A.* 109, 13052–13057
20. Miotto, O. *et al.* (2013) Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat. Genet.* 45, 648–655
21. Imamura, H. *et al.* (2016) Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. *eLife* 5, e126
22. Downing, T. *et al.* (2011) Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* 21, 2143–2156
23. Holt, K.E. *et al.* (2008) High-throughput sequencing provides insights into genome variation and evolution in *Salmonella typhi*. *Nat. Genet.* 40, 987–993
24. Doyle, S.R. *et al.* (2017) Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. *PLoS Negl. Trop. Dis.* Published online July 26, 2017. http://dx.doi.org/10.1371/journal.pntd.0005816
25. Choi, Y.-J. *et al.* (2017) Genomic introgression mapping of field-derived multiple-anthelmintic resistance in *Teladorsagia circumcincta*. *PLoS Genet.* 13, e1006857
26. Bourguinat, C. *et al.* (2015) Macrocyclic lactone resistance in *Dirofilaria immitis*: failure of heartworm preventives and investigation of genetic markers for resistance. *Vet. Parasitol.* 210, 167–178
27. Doyle, S.R. *et al.* (2018) A major locus for ivermectin resistance in a parasitic nematode. *bioRxiv* Published online April 23, 2018. http://dx.doi.org/10.1101/298901
28. Zamanian, M. *et al.* (2018) Discovery of genomic intervals that underlie nematode responses to benzimidazoles. *PLoS Negl. Trop. Dis.* 12, e0006368
29. Sallé, G. *et al.* (2018) The global diversity of a major parasitic nematode is shaped by human intervention and climatic adaptation. *bioRxiv* https://doi.org/10.1101/450692, Posted October 24, 2018
30. Howe, K.L. *et al.* (2017) WormBase ParaSite – a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10
31. Aurrecoechea, C. *et al.* (2017) EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Res.* 45, D581–D591
32. Davidse, L.C. and Flach, W. (1977) Differential binding of methyl benzimidazol-2-yl carbamate to fungal tubulin as a mechanism of resistance to this antimitotic agent in mutant strains of *Aspergillus nidulans*. *J. Cell Biol.* 72, 174–193

33. van Tuyl, J.M. (1977). *Genetics of Fungal Resistance to Systemic Fungicides*, Dissertation, Landbouwhogeschool Wageningen. http://library.wur.nl/WebQuery/wurpubs/71131
34. Driscoll, M. *et al.* (1989) Genetic and molecular analysis of a *Caenorhabditis elegans* beta-tubulin that conveys benzimidazole sensitivity. *J. Cell Biol.* 109, 2993–3003
35. Ghisi, M. *et al.* (2007) Phenotyping and genotyping of *Haemonchus contortus* isolates reveals a new putative candidate mutation for benzimidazole resistance in nematodes. *Vet. Parasitol.* 144, 313–320
36. Kwa, M.S. *et al.* (1994) Benzimidazole resistance in *Haemonchus contortus* is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1. *Mol. Biochem. Parasitol.* 63, 299–303
37. Rezansoff, A.M. *et al.* (2016) Evidence from two independent backcross experiments supports genetic linkage of microsatellite Hcms8a20, but not other candidate loci, to a major ivermectin resistance locus in *Haemonchus contortus*. *Int. J. Parasitol.* 46, 653–661
38. Laing, R. *et al.* (2016) Analysis of putative resistance gene loci in UK field populations of *Haemonchus contortus* after 6 years of macrocyclic lactone use. *Int. J. Parasitol.* 46, 621–630
39. Elmahalawy, S.T. *et al.* (2018) Genetic variants in dyf-7 validated by droplet digital PCR are not drivers for ivermectin resistance in *Haemonchus contortus*. *Int. J. Parasitol. Drugs Drug Resist.* 8, 278–286
40. Dent, J.A. *et al.* (2000) The genetics of ivermectin resistance in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 2674–2679
41. Gilleard, J.S. and Redman, E. (2016) Genetic diversity and population structure of *Haemonchus contortus*. *Adv. Parasitol.* 93, 31–68
42. Hahnel, S.R. *et al.* (2018) Extreme allelic heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.* 14, e1007226
43. Sham, P.C. and Purcell, S.M. (2014) Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15, 335–346
44. Korthauer, K. *et al.* (2018) A practical guide to methods controlling false discoveries in computational biology. *bioRxiv* Published online October 31, 2018. http://dx.doi.org/10.1101/458786
45. Visscher, P.M. *et al.* (2017) 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22
46. Coles, G.C. *et al.* (1992) World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. *Vet. Parasitol.* 44, 35–44
47. Awadzi, K. *et al.* (2004) An investigation of persistent microfilaridermias despite multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Ann. Trop. Med. Parasitol.* 98, 231–249
48. von Samson-Himmelstjerna, G. *et al.* (2009) Standardization of the egg hatch test for the detection of benzimidazole resistance in parasitic nematodes. *Parasitol. Res.* 105, 825–834
49. Evans, C.C. *et al.* (2013) Development of an *in vitro* bioassay for measuring susceptibility to macrocyclic lactone anthelmintics in *Dirofilaria immitis*. *Int. J. Parasitol. Drugs Drug Resist.* 3, 102–108
50. Lacey, E. *et al.* (1990) *A Larval Development Assay for the Simultaneous Detection of Broad Spectrum Anthelmintic Resistance.* https://publications.csiro.au/rpr/pub?list=BRO&pid=procite:29dfd251-2e10-4fb5-8eb6-c59ff48d98aa
51. Storey, B. *et al.* (2014) Utilization of computer processed high definition video imaging for measuring motility of microscopic nematode stages on a quantitative scale: 'The Worminator'. *Int. J. Parasitol. Drugs Drug Resist.* 4, 233–243
52. Jombart, T. *et al.* (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94
53. Price, A.L. *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909
54. Tzeng, J. *et al.* (2008) Multidimensional scaling for large genomic data sets. *BMC Bioinformatics* 9, 179
55. Yang, J. *et al.* (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82
56. Speed, D. *et al.* (2017) Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49, 986–992
57. Small, S.T. *et al.* (2016) Population genomics of the filarial nematode parasite *Wuchereria bancrofti* from mosquitoes. *Mol. Ecol.* 25, 1465–1477
58. Doyle, S.R. *et al.* (2018) A genome resequencing-based genetic map reveals the recombination landscape of an outbred parasitic nematode in the presence of polyploidy and polyandry. *Genome Biol. Evol.* 10, 396–409
59. McCarthy, M.I. *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369
60. Andersen, E.C. *et al.* (2015) A powerful new quantitative genetics platform, combining caenorhabditis elegans high-throughput fitness assays with a large collection of recombinant strains. *G3* 5, 911–920
61. Cook, D.E. *et al.* (2017) CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* 45, D650–D657
62. Ghosh, R. *et al.* (2012) Natural variation in a chloride channel subunit confers avermectin resistance in *C. elegans*. *Science* 335, 574–578
63. Pica-Mattoccia, L. *et al.* (2009) Genetic analysis of decreased praziquantel sensitivity in a laboratory strain of *Schistosoma mansoni*. *Acta Trop.* 111, 82–85
64. Cioli, D. *et al.* (1992) *Schistosoma mansoni*: hycanthone/oxamniquine resistance is controlled by a single autosomal recessive gene. *Exp. Parasitol.* 75, 425–432
65. Anderson, T.J.C. *et al.* (2018) Genetic crosses and linkage mapping in schistosome parasites. *Trends Parasitol.* 34, 982–996
66. Redman, E. *et al.* (2012) Introgression of ivermectin resistance genes into a susceptible *Haemonchus contortus* strain by multiple backcrossing. *PLoS Pathog.* 8, e1002534
67. Sargison, N.D. *et al.* (2018) A method for single pair mating in an obligate parasitic nematode. *Int. J. Parasitol.* 48, 159–165
68. Hunt, P.W. *et al.* (2009) The use of DNA markers to map anthelmintic resistance loci in an intraspecific cross of *Haemonchus contortus*. *Parasitology* 137, 705
69. Hodgkinson, J. *et al.* (2013) Identification of putative markers of triclabendazole resistance by a genome-wide analysis of genetically recombinant *Fasciola hepatica*. *Parasitology* 140, 1523–1533
70. Lynch, M. *et al.* (2014) Population-genetic inference from pooled-sequencing data. *Genome Biol. Evol.* 6, 1210–1218
71. Chevalier, F.D. *et al.* (2014) Efficient linkage mapping using exome capture and extreme QTL in schistosome parasites. *BMC Genomics* 15, 617
72. Guo, Y. *et al.* (2017) Networks underpinning symbiosis revealed through cross-species eQTL mapping. *Genetics* 206, 2175–2184
73. Martis, M.M. *et al.* (2017) RNA-Seq *de novo* assembly and differential transcriptome analysis of the nematode *Ascaridia galli* in relation to *in vivo* exposure to flubendazole. *PLoS One* 12, e0185182
74. Wit, J. and Gilleard, J.S. (2017) Resequencing helminth genomes for population and genetic studies. *Trends Parasitol.* 33, 388–399
75. Luo, X. *et al.* (2017) Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in *Haemonchus contortus*. *Parasites Vectors* 10, 31
76. Bourguinat, C. *et al.* (2017) Genetic profiles of ten *Dirofilaria immitis* isolates susceptible or resistant to macrocyclic lactone heartworm preventives. *Parasites Vectors* 10, 504

77. Avramenko, R.W. *et al.* (2017) The use of nemabiome metabarcoding to explore gastro-intestinal nematode species diversity and anthelmintic treatment effectiveness in beef calves. *Int. J. Parasitol.* 47, 893–902

78. Ali, Q. *et al.* (2018) Emergence and the spread of the F200Y benzimidazole resistance mutation in *Haemonchus contortus* and *Haemonchus placei* from buffalo and cattle. *bioRxiv* Published online September 25, 2018. http://dx.doi.org/10.1101/425660

79. Tydén, E. *et al.* (2014) Deep amplicon sequencing of preselected isolates of *Parascaris equorum* in β-tubulin codons associated with benzimidazole resistance in other nematodes. *Parasites Vectors* 7, 410