The global diversity of a major parasitic nematode is shaped by human intervention and climatic adaptation

Sallé G.[1,2,*], Doyle S.R.[1], Cortet J.[2], Cabaret J.[2], Berriman, M.[1], Holroyd N.[1], Cotton J.A[1]

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom

2. INRA - U. Tours, UMR 1282 ISP Infectiologie et Santé Publique, Centre de recherche Val de Loire, Nouzilly, France

* corresponding author

Abstract

The gastrointestinal parasite *Haemonchus contortus* is the most pathogenic nematode of veterinary interest and a model for many aspects of parasitic nematode biology. Using single-worm whole genome sequencing, we have performed the most exhaustive survey of genome-wide diversity undertaken for any parasitic nematode, considering 19 populations spanning five continents. Phylogenetic inference supports an African origin for the species. However, contemporary global distribution has been mediated by modern human movement, with evidence for parasites spreading during the transatlantic slave trade and colonisation of Australia presented. Like other veterinary and human parasitic nematodes, *H. contortus*

infections are controlled by widespread administration of anthelmintic drugs. Strong selection in response to benzimidazole treatment has impacted genetic diversity surrounding the β-tubulin locus independently throughout the world. Further, regions of high diversifying selection were enriched in genes involved in response to drugs as well as other anthelmintic-associated biological functions including pharyngeal pumping and oviposition. From these analyses, we identified genes putatively linked to ivermectin resistance that could be used in monitoring drug resistance. Finally, we describe genetic signatures of climate-driven adaptation, revealing that epigenetic regulation and components of the *dauer* pathway may play a role in adaptation in the face of climatic fluctuations. These results are the first to identify signatures of genetic adaptation to climate in a parasitic nematode, and provides insight into the ongoing expansion in the range of *Haemonchus contortus*, which may have consequences for the management of this parasite.

Keywords: *Haemonchus contortus*, helminth, parasitic nematode, global diversity, genomics, drug resistance, climate change

Introduction

Nematodes have evolved to exploit a wide diversity of ecological niches. While many sustain a free-living lifestyle, parasitic nematodes rely on one or more hosts to complete their life cycle. Many parasitic nematodes undergo a complex series of morphological changes, linked to migration through their hosts to establish a mature infection[1]. The complex life cycles of parasite species may involve both intermediate hosts, vectors or time spent in the environment, where they face harsh and variable conditions such as frost or drought that they must withstand between infection of their hosts[2]. Parasitic nematodes must thus adapt to a wide range of threats, including predation, climate and the immune responses of various host species. They have been enormously successful at this, with many thousands of nematode species infecting a great diversity of both plants and animals[3,4].

The evolutionary success of parasitic nematodes comes at a cost to humans, either directly as they significantly impact human health (amounting to a loss of 10 million disability-adjusted life-years)[3], or indirectly via major economic losses in plant[4] and livestock production[5], and parasite control. The control of parasitic nematodes of veterinary and medical importance relies almost exclusively on anthelmintic drugs, administered on a recurrent basis in livestock[6] and through mass-drug administration program in humans[7]. Although the success of such strategies was originally undeniable, the emergence of drug resistant parasites of veterinary interest[5], or the reported lack of efficacy in human-infective species[8], threatens ongoing control efforts for many parasitic infections. Vaccines offer an attractive alternate control strategy against these parasites: the extensive genetic diversity and immune-regulatory properties of parasites has however greatly hampered vaccine

development[9], and although two licensed vaccines are currently available for veterinary purposes[10], transcriptomic plasticity of the parasite following vaccine challenge may contribute to circumvent the vaccinal response of their host[10]. It is therefore clear that novel, sustainable control strategies are required. The potential of helminths to adapt to – and thus escape – control measures lies in their underlying genetic diversity. A greater understanding of the extent of this diversity and the processes that shape it throughout their range should provide insight into the mechanisms by which they adapt, and may identify new targets which may be exploited for control.

The trichostrongylid *Haemonchus contortus* is a gastrointestinal parasite of ruminants in tropical and temperate regions throughout the world, and causes significant economic and animal health burden particularly on sheep husbandry. It is also emerging as a model parasitic nematode system for functional and comparative genomics, largely due to its rapid ability to acquire drug resistance, the relative tractability of its life-cycle under laboratory conditions[11], the development of extensive genomic resources[12,13], and its relatively close relationship with other clade V parasitic nematodes of both veterinary and medical importance (i.e, human hookworms and other GIN of livestock)[14]. Exploiting an existing collection of animal parasitic nematodes[15], we have used whole genome sequencing of 223 individual *H. contortus* collected from sheep or goats in 12 countries spanning five continents, to characterise genome- and population-wide genetic diversity throughout its range. To our knowledge, this is the most exhaustive survey of genome-wide diversity undertaken for any parasitic nematode, unravelling old and new genetic connectivity

influenced by human history, and signatures of selection in response to anthelmintic exposure and local climatic variation.

Results

Haemonchus contortus populations are genetically diverse, with large effective population sizes

Whole genome sequencing of 223 individuals from 19 populations (Fig. 1a; Supplementary Table 1) identified 23,868,644 SNPs with an average mean genome-wide distribution of 1 SNP per 9.94 bp. Only a limited fraction of these markers was shared across populations (n = 411,574 SNPs common across more than half the individuals with MAF>5%). Estimates of nucleotide diversity (π) in populations with at least five individuals ranged from 0.44% in one population from South-Africa (STA.2) to 1.3% in the Namibian population (NAM; Supplementary Table 2). Variance in π across autosomes among populations was partly explained by population mean coverage ($F_{(1,84)}$ = 9.49, P = 0.003; Supplementary note). To account for this bias, estimates were obtained from three subsets of populations with the highest coverage (greater than 8× on average) from France (FRA.1 and FRA.2), Guadeloupe (FRG) and Namibia (NAM) that yielded slightly higher values ranging between 0.65 and 1.14%. Overall, these data were roughly equivalent with diversity levels of *Drosophila melanogaster* (ranging between 0.53% and 1.71%)[16] and represented approximately five- to ten-fold greater diversity than human species[17] (0.119% across African, American and Asian populations; Supplementary Fig. 1).

To begin to explore the global diversity of *H. contortus*, we performed a principal component analysis (PCA) of genetic variation, revealing three broad genetic clusters (Fig. 1b) distinguishing: (i) Sub-tropical African populations (NAM, STA and ZAI), (ii) Atlantic populations including Morocco (MOR), São Tomé (STO), Benin

(BEN), Brazil (BRA) and Guadeloupe (FRG), and (iii) the remainder from the Mediterranean area (FRA, ACO) and Oceania (AUS.1, AUS.2 and IND). The greatest diversity among samples was identified in the African and South American populations, with East African samples spread along PC1 and West African and South American samples along PC2.

From the nucleotide diversity estimates, we inferred the current effective population size (N_e) of *H. contortus* to be between 0.60 and 1.05 million individuals, using the *C. elegans* [18] mutation rate and assuming a balanced sex ratio[19]. MSMC analysis, which models past recombination events based on heterozygosity patterns along the genome, revealed N_e has remained within this range of values for most of the sampled time interval, i.e. from 2.5 to 100 thousand years ago (kya), with extreme estimates falling between 0.36 and 3.39 million individuals for FRA.1 (2.32 kya) and NAM (2.23 kya) populations respectively (Fig. 1c). Similar trends in recent population sizes were observed for all populations except the French population, for which N_e was highly variable and peaked at 4.22 million individuals approximately 700 years ago (Fig. 1c).

Global population connectivity of Haemonchus contortus is characterised by old and new migration

To explore the global connectivity between populations, we used a number of complementary approaches. Phylogenetic connectivity determined by nuclear (Supplementary Fig. 2) and mitochondrial (Fig. 2a; PCA of mtDNA genetic diversity is presented in Supplementary Fig. 3) diversity broadly supported the initial PCA analysis (Fig. 1b), also revealing the three main groups of samples partitioned into

broad geographic regions. However, the presence of close genetic relationships between geographically distant populations, resulting in a weak phylogeographic signal (Mantel's test $r = 0.10$, $P = 0.001$), was inconsistent with a simple isolation-by-distance scenario. This observation was supported by, for example, little genetic structure between French (FRA) and Oceanian (AUS.1, AUS.2, IND) populations measured by F_{ST} (Fig. 2b). The greatest genetic dissimilarity was found between subtropical African populations and the French and Australian populations, with mean F_{ST} across comparisons of 0.21 (F_{ST} range according to population pairs: 0.06 - 0.42) and 0.24 (F_{ST} range: 0.17 - 0.33) respectively (Fig. 2b, Supplementary Fig. 4). The divergence of African populations was likely due to the higher within-population diversity (+14% higher mitochondrial nucleotide diversity) relative to others ($P = 0.004$; mitochondrial nucleotide diversity of 0.63% ± 0.37%, 0.58% ± 0.31%, 0.66 ± 0.32%, 0.6% ± 0.35%, for Mediterranean, Oceanian, American and south-African populations respectively).

The higher genetic differentiation (8.9% difference, $F_{(1,134)} = 11.36$, $P = 0.0009$) and higher genetic dissimilarity (5.7% difference, $F_{(1,7567)} = 581$, $P < 10^{-4}$) between STO samples and other populations relative to every other pairwise comparisons certainly reflects isolation of worms on the island from continental populations (Fig. 2b). However, phylogenetic analysis revealed a main ancestral lineage formed by BEN and STO worms, from which additional divergence occurred in a stepwise manner to give rise to an Oceanian clade, a subtropical African haplogroup and a Mediterranean cluster (Fig. 2a). Contribution of the BEN and STO ancestry to European and Australian was also evident by the admixture pattern of French and one Australian populations (Fig. 2c).

FRG samples revealed a mixed ancestry from West African and Mediterranean populations. A subset of FRG samples showed limited genetic differentiation to FRA populations (F_{ST} range = 0.07 - 0.10; Fig. 2b), whereas the remaining FRG samples were genetically similar to populations from the West African coast, *i.e.* ACO, BEN, MOR, STO (Fig. 2b & c), as indicated by lower F_{ST} estimates with these populations (F_{ST} range = 0.07 - 0.13; Wilcoxon's test, P = 0.07; Fig. 2b). Interestingly, a close relationship was identified between STO and FRG (F_{ST} = 0.10; Fig. 2b), which was supported by admixture analysis (Fig. 2c, Supplementary Fig. 5 and 6) that suggested that BEN and STO populations share a genetic ancestry with North African populations and American populations, i.e. FRG and BRA. This conflicting genetic origin among FRG sub-populations would support the higher nucleotide diversity observed in FRG as a whole (Supplementary Table 2), and may also support the inflection in N_e over time seen specifically in this population (Fig. 1c).

The patterns of genetic connectivity support at least three distinct migration events in time and space. First, we detect an out-of-Africa scenario, whereby the greatest diversity was sampled within Africa, and that populations outside of Africa represent a subset of this diversity. Consistent with this hypothesis, nuclear genome variation of non-African populations experienced a genetic bottleneck that occurred between 2.5 and 25 kya that is not present in the African populations sampled (Fig. 1c). Bayesian coalescent estimate of this initial divergence from mitochondrial genome data yielded an overlapping time range of between 3.6 kya and 4.1 kya (Supplementary Fig. 7). Secondly, the genetic connectivity of West African and population from the Americas was consistent with spreading of parasites during the

trans-Atlantic slave trade movement. This was particularly evident from FRG populations that had been affected by the contribution of colonization and slave trade that occurred in the West Indies under French influence[20]. The third pattern of connectivity likely reflects British colonisation of Australia in late 1700's: the interweaving of Australian and South-African worms into the Mediterranean phylogenetic haplogroup (Fig. 2b) may mirror the foundation of Australian Merino sheep, which were first introduced into Australia from South Africa, before additional contributions from Europe and America were made[21,22]. The admixture pattern observed for worms from the two countries matched their shared ancestry (Fig. 2c), as well as a genetic connection between America and Australia (seen for the AUS.1 population; Fig. 2c). These complex patterns reiterate that human movement has played an important role in shaping the diversity of this livestock parasite throughout the world.

Anthelmintic resistance has left distinct patterns of diversity in the Haemonchus contortus genome

The widespread use of anthelmintics has been and remains the primary means of *H. contortus* control worldwide. Extensive use of anthelmintics has resulted in the independent emergence of drug-resistant populations through the world, which now limits farming in some areas. Strong selection should impact the distribution of genetic variation in populations; signatures of selection may reveal drug-resistance associated genes, knowledge of which may contribute to the monitoring the spread of drug resistant populations and the design of new control strategies. We

specifically looked for evidence of selection in populations phenotypically characterised to be resistant to benzimidazoles and ivermectin.

The genetic determinants of benzimidazole resistance is perhaps the best characterised of all anthelmintics, with three residue changes in the beta tubulin isotype 1 (*Hco-tbb-iso-1)* locus strongly associated with phenotypic resistance. The region surrounding the *Hco-tbb-iso-1* locus showed indications of a selective sweep in the resistant populations. Focusing on three populations with highest coverage, we found an average 2.31-fold reduction of Tajima's D coefficient within 1 Mbp (Fig. 3a) and an average 33% reduction in nucleotide diversity (Supplementary Fig. 8) within this region relative to the rest of chromosome I. This signature was most evident in the FRG and NAM populations, but was weaker in French worms (FRA.1) due to both phenotypically susceptible and resistant individuals being present (Supplementary Table 3). Phased genotypic information over the whole *Hco-tbb-iso-1* (Supplementary Fig. 9) revealed that Australian and African populations had the same level of divergence at this locus as found genome-wide in both nuclear and mitochondrial genomes, suggesting that resistant mutation(s) occurred independently on different genetic backgrounds. An exception was between French and Guadeloupian individuals, where little divergence in their haplotypes may relate to an introgression event from mainland France into the West Indies (Supplementary Fig. 9). A topology analysis run over a 20 Kbp region spanning the *Hco-tbb-iso-1* locus supported supported the shared phylogenetic origin between FRA.1 and FRG (topology 3), whereas the surrounding region was in favour of topologies congruent with overall population structure (Fig. 3b). This finding holds when considering either São Tomé (Fig. 3b) or Moroccan populations as a population with close ancestry with FRG (Supplementary Fig. 10).

We analysed the presence of the three well-characterised resistance-associated mutations affecting codon positions 167, 198 and 200 (Fig. 3c; Supplementary Table 3; Supplementary Table 4; Supplementary Fig. 11). The F200Y homozygous genotypes were most common and widespread resistant genotype (n = 39; T/T-A/A-A/A), and accounted for all samples from the Guadeloupe (n = 14; T/T-A/A-A/A) and the White-River South-African (STA.3; n = 6; T/T-A/A-A/A) populations. Variants at codons 167 and 198 were much less common, i.e. 13 mutant allelic carriers were observed at position 198 and only one F167Y mutant was identified in a French population. No double homozygous mutants were found, but three individuals from France, Australia and South-Africa were heterozygous at both positions 198 and 200. In each case, inspection of the sequencing reads revealed that the two mutations never appeared together in the same sequencing read, suggesting they cannot co-occur in cis (on the same chromosome copy). Genotypes were consistent with population-level benzimidazole efficacies as measured by the percentage of egg excretion after treatment (Supplementary Table 4). Samples analysed from suspected susceptible populations always presented with the susceptible genotype (n = 20; T/T-A/A-T/T).

The genetic determinants of ivermectin resistance are largely unknown, but many genes have been proposed to be associated with resistance; although one interpretation of this observation is that ivermectin resistance is a multigenic trait, a major locus associated with resistance in Australian and South African *H. contortus* has recently been mapped to chromosome V[23]. To examine the effect of ivermectin resistance in our data, a pairwise differentiation scan was performed between known

ivermectin-resistant populations and other populations sharing same genetic ancestry, *i.e.* STA.1 and STA.3 versus NAM and ZAI in Africa, AUS.1 against AUS.2 in Australia (Fig. 3d). The previously identified QTL region on chromosome V[23] appeared differentiated, particularly between pairwise comparisons involving the South-African resistant population (Fig. 3d). A second major differentiation hotspot spanned a 3-Mbp region of chromosome I and encompassed the β-tubulin isotype 1 (*Hco-tbb-iso-1*) locus (Fig. 3d). Although non-synonymous mutations at codon positions 167, 198 and 200 of the β-tubulin isotype 1 are associated with resistance to benzimidazoles, it has been proposed that there may also be an association between this gene and ivermectin resistance[24,25]. Although our data seem to support this association, an attempt to narrow-down the differentiation signal in this region found that significant differentiation in only two 10-kbp windows in two comparisons (NAM vs STA.1 and NAM vs. STA.3, and these two windows did not overlap the *Hco-tbb-iso-1* locus. Furthermore, the fact that all of our ivermectin-resistant populations were also benzimidazole-resistant suggests that this signal is confounded; pairwise F_{ST} analyses between ivermectin-resistant and -susceptible populations from the field will always be biased toward strong differentiation around the *Hco-tbb-iso-1* locus due to loss of diversity in benzimidazole-resistant populations. As such, it was not possible to confirm the putative association between β-tubulin isotype 1 and ivermectin resistance. We note that the lack of QTL evidence in this region from controlled genetic crosses using ivermectin selection[23] (Doyle, personal communication) supports the conclusion that standing genetic variation at the β-tubulin locus is unlikely to be directly influenced by ivermectin.

To further investigate more subtle signatures of selection, we computed the XP-CLR coefficient, which simultaneously exploits within-population departure from neutrality and between-population allele-frequency differences[26]. This analysis relies on called genotypes rather than genotype likelihoods, but is robust to SNP uncertainty. Within continent pairwise comparisons of African (NAM, MOR, STA.1, STA.3, ZAI) and Australian populations (AUS.1, AUS.2) yielded 1,740 hotspots of diversification, 48% of which were contained within a gene locus (Supplementary Fig. 14). Among these hotspots, two known candidate genes associated with ivermectin resistance were identified, namely an ivermectin sensitive glutamate-gated chloride channel[27] (*HCOI00617300*, *glc-4* ortholog) on chromosome I, and a P-glycoprotein coding gene already involved in ivermectin susceptibility in the equine ascarid *P.* equorum[28], (*HCOI00233200, pgp-11* ortholog) on chromosome V. While the former showed significant selection between the only two populations with confirmed ivermectin-resistance (AUS.1) and -susceptibility (AUS.2), the latter was found to be significant after averaging across all pairwise comparisons, suggesting its role may not be specifically related to ivermectin resistance. GO term enrichment analysis of all XP-CLR significant regions identified 26 significant terms (Supplementary Table 5). The top ten most significant GO terms encompassed enzymatic-related activity, neurotransmitter transporter activity (GO:0005326, $P = 5.8 \times 10^{-4}$) and response to drug (GO:0042493, $P = 4.3 \times 10^{-3}$). Additional significant biological process associated terms were related to phenotypes tightly linked to anthelmintic effects. For example, pharyngeal pumping[29] (GO:0043051, $P = 5.8 \times 10^{-3}$) and oviposition (GO:0046662, $P = 8.6 \times 10^{-3}$) genes were enriched, both of which are phenotypes linked to ivermectin and its effect on parasite fecundity[30]. Neurotransmission is the primary target of anthelmintics such as macrocyclic lactones[31] and levamisole[32];

genes with GO terms related to neurotransmission (GO:0001505, $P = 9.1 \times 10^{-3}$) were significantly enriched across comparisons. Anthelmintic-associated GO terms including “response to drug”, “regulation of neurotransmitter” and “neurotransmitter transporter activity” were significantly enriched in every comparison involving a South-African multi-resistant field isolates (STA.1, Supplementary Table 6). Four genes, *HCOI00389600*, *HCOI00032800*, *HCOI00243900*, *HCOI00489500*, thus seem candidates for drug resistance loci in *H. contortus*, as supported by the functions of their *C. elegans* orthologs (*snf-9*, *unc-24*, *B0361.4*, *aex-3* respectively). The *snf-9* gene encodes a neurotransmitter:sodium symporter belonging to a family of proteins involved in neurotransmitter reuptake, and the latter three genes are expressed in neurons. *Unc-24* and *B0361.4* are involved in response to lipophilic compounds and *aex-3* is critical in synaptic vesicle release[33].

Anthelmintic treatments appear to be a major selective constraint for *H. contortus* populations. Several positional candidates underpinning synaptic function, namely *snf-9*, *aex-3* and *glc-4*, were likely associated with ivermectin resistance and a distinct selection footprint was found at the β-tubulin locus, but it is unlikely that this locus contributes to ivermectin-resistance.

Climatic adaptation has shaped genomic variation between populations

In addition to putative anthelmintic-related GO terms, response to stress (GO:0006950, $P = 5.9 \times 10^{-3}$) was among the top ten significant biological process ontologies associated with genes under diversifying selection. Although anthelmintic exposure would be associated with significant stress on susceptible (and perhaps resistant or tolerant) parasites, all free-living stages of *H. contortus* will be exposed to and must tolerate abiotic factors such as temperature or humidity prior to infection

of a new host. To evaluate the impact of such climatic stressors, a genome scan for genetic differentiation between populations categorised by their climatic conditions, i.e. arid (Namibia), temperate (France mainland) and tropical (Guadeloupe), was performed (Fig. 4a). A major signal of differentiation formed of two windows (6.925 – 6.995 Mbp and 7.165 - 7.205 Mbp on chromosome I) was shared by all pairwise comparisons. Six genes were found within these two regions (Supplementary Table 7), among which orthologs of *C. elegans* genes *hprt-1*, *cpb-1, B0205.4* were identified. A highly differentiated 260-Kbp discontinuous region of chromosome III also repeatedly occurred across comparisons. A window between 31.155 Mbp and 31.185 Mbp was common to comparisons involving populations from arid areas (NAM vs. FRA.1 and NAM vs. FRG), and a second window (31.345 Mbp to 31.415 Mbp) was common to comparisons between tropical populations and others (FRG *vs*. FRA.1 and FRG *vs*. NAM). While the two annotated genes in the latter region were not associated with any biological description, the first window overlapped a chromo-domain containing protein (*HCOI01540500*; Supplementary Table 7). This gene is an ortholog of *Pc* (*FBgn0003042*) in *D. melanogaster*, a chromo-domain subunit of the Polycomb PRC1-complex that specifically recognizes trimethylated lysine7 of histone 3 (H3K27me3)[34]. None of the remaining differentiated windows appeared significantly differentiated across the three possible comparisons.

To further explore the impact of climatic conditions of genetic diversity, we used a gradient forest analysis to determine the relative impact of specific bioclimatic variables on genetic diversity. Bioclimatic variables are derived from monthly temperature and precipitation records and aims to represent annual trends or seasonality (Supplementary Fig. 14; Supplementary Table 8). This analysis builds

bootstrapped subsets of SNP frequencies to determine the climatic variable best able to partition allele frequencies in homogeneous groups and estimate the proportion of genetic variance explained by each variable. Annual precipitation (BIO12), and temperature annual range (BIO7) were revealed to be the most important bioclimatic variables impacting genetic variation (Fig. 4b). To identify genes that might be impacted by these variables, were performed an association analysis between each of these two environmental variables and the genome-wide data, after accounting for population structure between populations. In total, 17 and 25 significant associations (5% FDR; 49,370 SNPs tested) were found with BIO7 and BIO12 respectively (Fig. 4a bottom panel; Supplementary Table 8). Consistent with the initial differentiation scan, chromosomes I and III harboured most of the associations (n = 13 and n = 11, respectively). On chromosome I, eight (of the 13) associations with BIO12 fell within a 6 Kbp window (7,177,538 bp and 7,183,617 bp), overlapping the region universally differentiated between climatic areas. Chromosome III contained 11 significant associations for both BIO7 (n = 6) and BIO12 (n = 5). The two most significant associations were also found on this chromosome for BIO7 (positions 26,824,240 and 26,824,197 bp; $P = 6.34 \times 10^{-9}$ and $P = 1.39 \times 10^{-7}$, respectively). These two SNPs fell within the *HCOI00198200* locus, which codes for a metallopeptidase M1, an ortholog of the *C. elegans* gene *anp-1*. Additional BIO7-associated SNPs were found within (16,544,889 bp) and in the vicinity (15,905,141 & 15,905,191 bp) of a solute carrier coding gene (*HCOI00312500*) on chromosome IV, whose orthologs in *C. elegans* are under the control of *daf-12*, a key player in dauer formation. Of note, two *H. contortus* orthologs of components of the dauer pathway in *C. elegans*, namely *tax-4* (*HCOI00661100*) and *daf-36* (*HCOI00015800*), were among genes under diversifying selection

(Supplementary Fig. 14). These genes were, however, not identified in the association analyses against bioclimatic variables.

Discussion

Recent advances in genomic resources[13,23] and dissection of its adaptation to anthelmintic drug[23] or vaccine[10] exposure have increased the utility of *H. contortus* as a model system to study anthelmintic resistance, and toward developing more sustainable and integrated management practices for gastrointestinal nematodes. By exploiting a broad collection of samples from globally-distributed populations, we have performed an in-depth characterization of *H. contortus* genetic diversity, and identified important drivers shaping the genome of this parasite. *H. contortus* populations displayed high levels of nucleotide diversity, consistent with early estimates based on mitochondrial data[35], and recent re-sequencing experiments of inbred isolates[12,23]. These extreme levels of genetic diversity were also reported in another free living nematode species[36], and are thought to arise from both a large census population[37] and the high fecundity of *H. contortus* females[19].

Major past human migrations and associated sheep movements have contributed to the mixing of parasite populations, which partly accounts for the limited genetic structure and extensive admixture between some of our globally distributed populations. Our data supported an out of Africa expansion derived from ancestral populations from Western Africa, with a bottleneck dated back between 2.5 to 25 kya. Sheep domestication originally took place in the Middle East around 10 kya[38], before introduction into Eastern Africa and subsequent spread toward southern Africa they reached between 2 and 2.5 kya likely through cultural diffusion [39]. Our sampling did not cover these Eastern regions and could not resolve the exact origin of *H. contortus* populations. However, the timing of the *H. contortus* population bottleneck is compatible with major migrations of pastoralist populations that

ultimately resulted in the import of small ruminants into Southern Africa through both central and Namibian routes[39]. The simultaneous increase in rainfall in central Africa around 10.5 kya [40] would have supported the multiplication and dispersal of *H. contortus*. In addition, early radiation towards Asia observed in our data was congruent with evidence obtained from the study of retroviral insertions within the sheep genome that suggested direct migration between Africa and southwest Asia[41], and with the timing of Asian sheep expansion between 1.8 and 14 kya [42].

The co-transportation of livestock, including sheep, during discovery and colonization of America resulted in an admixed Caribbean sheep population[43,44]. Woolly Churra sheep were originally brought to the Caribbean by Spanish conquistadors[44,45], before West African breeds more suited to tropical conditions were later shipped along with enslaved people during the slave trade[44,45]. Congruent with this history, the genetic diversity of *H. contortus* populations in Guadeloupe displayed both a Mediterranean ancestry while being tightly linked to São Tomé and Benin populations. This latter link with tropical Western Africa recalls that Benin was the major harbour for slave trade to the French West Indies[20], and therefore we speculate that the slave trade was a major means of dispersal of worms between these regions. The Mediterranean ancestry of *H. contortus* populations in Guadeloupe and the close relationship with French worm populations suggest additional sheep transport from France occurred. While these movements are difficult to track, live sheep were shipped aboard slave trade vessels departing from French harbors[46], and may have introduced European *H. contortus* to the island. Based on historical data[44,45], a Spanish lineage of *H. contortus* on Guadeloupe would also be expected to exist, though data are missing to confirm this. It can be speculated that the shared

ancestry between Guadeloupian and Moroccan worm populations might result from the introduction of Spanish Churra sheep, as this breed emerged in Spain while the region was under Arab influence between the 8th and 13th century[44,47]. Introduction of *H. contortus*-infected sheep from the Maghreb may have been associated with the spreading of African worms in Spain, and ultimately, Guadeloupe.

Adaptation to anthelmintic drugs was a major selective force shaping standing genetic variation across the populations analysed, underscoring the impact of regular drenching programs applied in the field worldwide. The β-tubulin coding gene provides a striking illustration of strong loss of diversity due to benzimidazole anthelmintics in many individual and distant populations. The co-occurrence of resistant haplotypes in disconnected populations supports the conclusion that resistance has appeared in several backgrounds independently[48]. However, shared haplotypes between France and Guadeloupe suggests resistant individuals were also transported. This finding emphasises the need for monitoring of parasitic populations during livestock trade, but also the risk of evolution of resistance in other parasitic nematodes, for example, parasites of medical interest that are treated with benzimidazoles in mass-drug administration programs[49].

In addition to benzimidazoles, ivermectin is an important anthelmintic for parasite control in both veterinary and medical settings[50]. Worldwide emergence of ivermectin-resistant veterinary parasitic nematodes[5] and evidence of reduced efficacy in human filarial nematodes [51] underline the urgency and importance of a better understanding of the mechanisms at play[50]. Our data identified genes previously associated with ivermectin resistance in parasitic nematodes of either

veterinary (*pgp-11 in Parascaris sp.*[28]) or medical (*aex-3* in *Onchocerca volvulus*[52]) interest that may be useful markers in the field to monitor drug efficacy. Of note, support for a major QTL for ivermectin resistance on chromosome V recently identified in an introgression experiment in two geographically and genetically diverse resistance isolates[23] was also present in our genome-wide data and warrants further investigation.

Our broad sampling throughout the global range of *H. contortus* has enabled the first analyses into climate-driven adaptation in a parasitic nematode. Understanding how climate shapes parasite genetic variation is of primary importance to foresee consequences of climate change on the parasites life cycle and ability to disperse. Our results suggest that adaptation toward annual precipitation was mostly under the control of variation on chromosome I, whereas a region on chromosome III was associated with both temperature- and precipitation-related variables. No obvious link with climatic adaptation could be inferred from current knowledge on chromosome I genes. In addition, the *Hco-tbb-iso-1* locus was close to the region of interest, and linked variation in allelic frequency at this locus as a result of benzimidazole selection cannot be ruled out. However, genes identified within the chromosome III region appeared more biologically relevant. First, the strongest genetic associations with annual temperature range were found in a metallopeptidase with zinc ion binding function (*anp-1* ortholog), an enzyme linked to drought stress tolerance in *Drosophila* [53]. Second, a Polycomb gene ortholog displayed strong genetic differentiation between arid- and wetter temperate or tropical environmental conditions; this gene has been linked to putative epigenetic regulation of xeric adaptation in *Drosophila melanogaster* whereby Polycomb

mutants display lower resistance to desiccation stress[54]. This finding is also corroborated by observations in plants showing Polycomb-mediated regulation of climate-induced phenotypes[55]. Further links to climatic adaptation include identification members of the dauer pathway[56]; *tax-4* and *daf-36* orthologs were under diversifying selection and a *daf-12*-respondent solute carrier was associated with annual temperature range. Dauer is a developmentally arrested stage in *C. elegans* that is triggered by environmental stress and mediates tolerance to unfavourable conditions until better conditions are met[56], and can occur in parasitic nematodes like *H. contortus* under semi-arid conditions[57]. These two findings have major implications for the sustainable control of parasitic nematodes in veterinary and medical settings. First, adaptation to climate change will be constrained by available genetic variability at the temperature-selected loci, hence possibly reducing parasite expansion in some regions. Second, better understanding of the relationship between climatic conditions and hypobiosis is critical to optimize treatment decision. Third, epigenetically-mediated plasticity may result in broad transcriptional regulations that may affect drug-resistance associated genes and might also affect expression patterns of subsequent generations. Characterizing these interactions further may help prevent the emergence of drug resistant populations.

In summary, our data describes the extensive global and genome-wide diversity of one of the most pathogenic parasitic nematodes, and how this diversity has been shaped by adaptation to its environment and evasion of drug exposure. The dispersal of *H. contortus* throughout the world has been shaped in large part by human movement, influenced by within Africa migration and more contemporary transport consistent with the slave trade from Africa to the Americas and colonization

of Australia. Our data highlight the impact of drug exposure and the evolution of resistance has on the genome, and we identify a number of genes that may be involved in tolerating ivermectin exposure that may be used to monitor drug resistance. Finally, we report the first evidence of genetic signatures of climate-driven adaptation, suggesting histone methylation marks and dauer pathway may play a role in adaptation to arid conditions. Understanding the mechanism(s) by which parasites tolerate and adapt in response to fluctuating environmental conditions will have major implications for field management of parasitic nematodes in both veterinary and medical settings. Further characterisation of these putative strategies, together with genetic covariation of drug resistance genes, should contribute to refining epidemiological models and guide treatment decision-trees for a more sustainable management of worm populations in the face of a changing climate.

Legends to Figures

Figure 1. Global diversity of Haemonchus contortus.

(a) Map depicts the global distribution of the populations sampled, coloured by geographical region (sand: South-America, brown: Western-Africa, dark green: Mediterranean area, black: Subtropical Africa, red: Australia). Shape indicates population resistance status to anthelmintics, either susceptible to fenbendazole and ivermectin (circle) or susceptible (square) to ivermectin and resistant, or resistant to fenbendazole and ivermectin (triangle).

(b) Principal component analysis based on genotype likelihood inferred from whole genome sequences of 223 individual males (1,638,715 variants considered). Samples are coloured by geographic region described in panel 1a.

(c) MSMC effective population size across time for 6 populations. Coloured shaded area represents range of values estimated from a cross-validation procedure with five replicates, computed by leaving one chromosome out at a time.

Figure 2. Global connectivity of Haemonchus contortus populations.

(a) Maximum likelihood phylogenetic tree based on 223 individual mitochondrial genomes, rooted on *Teladorsagia circumcincta* and *Haemonchus placei*. Circles indicate bootstrap support for each branch, blue if support was higher than 50%, red elsewhere. The root node has a 84% support. Branches leading to a sample identifier are coloured by geographic region described in Figure 1a.

(b) Matrix showing pairwise dissimilarity between individuals (upper right) and pairwise F_{ST} between populations (lower left).

(c) Admixture analysis for K = 3 clusters and for 223 individuals. Populations are presented sorted along their longitudinal range, and samples sorted by assignment to the 3 clusters.

Figure 3. Anthelmintic-mediated selection is a major driver of genetic variation

(**a**) Windowed Tajima's D estimated across a 10-Mbp window of chromosome I (pink) and within 1 Mbp surrounding *Hco-tbb-iso-1* (blue) for French, Guadeloupian and Namibian populations.

(**b**) Topology weighing analysis of a 20-Kbp window centred on *Hco-tbb-iso-1*. At each position, the weight of each of the three possible topologies inferred from 50 Kbp-windows are overlaid. Topology 2 (blue) corresponds to an isolation-by-distance history, while topology 3 (brown) would agree with shared introgressed material between worm populations from French mainland into Guadeloupe. The *Hco-tbb-iso-1* locus is indicated by vertical dashed lines.

(**c**) Genome-wide differentiation scan between pairs of ivermectin-resistant (AUS.1, STA.1, STA.3) and populations of unknown status. Chromosomes are coloured and ordered by name, i.e. from I (pink) to V (forest green). Horizontal line represents the 0.5% F_{ST} quantile cut-off. Vertical line on chromosome I points at *Hco-tbb-iso-1* locus.

Materials and methods

Sample DNA extraction and sequencing

A total of 267 individual male *H. contortus* were obtained from a collection held at INRA[16]. Samples were gathered between 1995 and 2011 (Supplementary Table 1), and stored in liquid nitrogen upon collection. Samples spanned 19 different populations across 12 countries (metadata for all samples is presented in detail in Supplementary Table 1). Three populations were ivermectin-resistant (AUS.1, STA.1, STA.3), and four populations were fenbendazole susceptible (Fig. 1; triangles; FRA.2, FRA.4, STO, ZAI).

DNA was extracted with the NucleoSpin Tissue XS kit (Macherey-Nagel GmbH&Co, France) following the manufacturer's instruction. Sequencing libraries were prepared as previously described[23]. DNA libraries were sequenced with 125 bp paired-end reads on an Illumina Hiseq2500 platform using V4 chemistry (Supplementary Table 1). A second round of sequencing was performed with 75 bp paired-end reads to increase the coverage of 43 samples (Supporting Fig. 15). After sequencing, two samples were identified to be heavily contaminated by kraken-0.10.6-a2d113dc8f[58] and were discarded. In total, 18 sequencing lanes consisting of 4,152,170,256 reads were sequenced, the raw data of which are archived under the ENA study accession PRJEB9837.

Sequencing data processing

Read mapping to both the mitochondrial and nuclear genomes (v3.0, available at ftp://ngs.sanger.ac.uk/production/pathogens/Haemonchus_contortus) was performed

using SMALT (http://www.sanger.ac.uk/science/tools/smalt-0) with a median insert size of 500 bp, k-mer length of 13 bp, and a stringency of 90%. For samples that had two or more BAM files (when split across multiple sequencing lanes), the BAM files were merged using samtools v.0.1.19-44428[59], and duplicated reads removed using Picard v.2_14_0 (https://github.com/broadinstitute/picard) before performing realignment around indels using Genome Analysis Toolkit (GATK v3.6)[60] RealignerTargetCreator. Mean coverage of the mitochondrial and genomic genomes were estimated using GATK DepthOfCoverage, revealing a lower than the estimated target coverage (original 8x) for most samples. Individuals with more than 80% of their mitochondrial sequence with at least 15 reads and a mean mitochondrial genome coverage of at least 20× were retained for population genetic inferences (n = 223 individuals).

Genomic SNP calling

To call SNPs, we used GATK HaplotypeCaller in GVCF mode, followed by joint genotyping across samples (GenotypeGVCFs) and extraction of variants (SelectVariants), resulting in a total of 30,040,159 unfiltered SNPs across the five autosomes. Sex determination in *H. contortus* is based on an XX/XO system; as only male worms were sequenced, their hemizygous X chromosome would have revealed limited phylogenetic information relative to the autosomes, and was henceforth excluded from further analysis.

Low coverage sequencing will inadvertently bias allele sampling at heterozygous SNPs, resulting in excess homozygous genotypes particularly if stringent filtering is applied during SNP calling. To circumvent this issue, we applied the GATK Variant Quality Score Recalibration

(https://gatkforums.broadinstitute.org/gatk/discussion/39/variant-quality-score-recalibration-vqsr), which first uses a reference (truth) SNP set to estimate the covariance between called SNP quality score annotations and SNP probabilities, followed by application of these probabilities to the raw SNPs of interest.

The reference “truth” SNP database was generated from the intersection of variants called from samples with at least a mean of 10× coverage (n = 13) using three independent SNP callers: (i) samtools mpileup (-q20 –Q20 –C50 -uD), (ii) Freebayes[61] v.9.9.2-13-gad718f9-dirty (--min-mapping-quality 20 --min-alternate-count 5 --no-indels --min-alternate-qsum 40 --pvar 0.0001 --use-mapping-quality --posterior-integration-limits 1,3 --genotype-variant-threshold 4 --use-mapping-quality --site-selection-max-iterations 3 --genotyping-max-iterations 25 --max-complex-gap 3), and (iii) GATK HaplotypeCaller followed by hard filtering (--QD<2, --DP>10000, --FS>60, --MQ<40, --MQRS <-12.5, --RPRS<-8). The three variant call sets were merged (GATK CombineVariants with --genotypeMergeOptions UNIQUIFY), resulting in an intersecting set of 794,606 SNPs (extracted with GATK SelectVariants). The GATK VariantRecalibrator model was trained with this reference SNP database with a 90% prior likelihood, before being subsequently applied to the raw set of SNPs (n = 30,040,159). The estimation was run for several truth sensitivity threshold values ranging from 90 to 99.9%. After visual inspection of the additional number of SNPs by using sensitivity tranche curves (Supplementary Fig. 16a), a 97% sensitivity threshold was applied to the raw SNP set with GATK ApplyRecalibration, resulting in a total set of 23,868,644 SNPs spanning the five autosomes. Variant depth of coverage (DP) and strand bias (FS) were the main drivers of SNP removal (Supplementary Fig. 16b).

Called SNPs were used for particular analyses (*Ne* trajectory through time, cross-population composite likelihood-ratio) that could not be performed under the probabilistic framework that relied on genotype likelihoods as implemented in ANGSD[62] v. 0.919-20-gb988fab (Supplementary note). These data were also used to estimate average differentiation between populations. However, within population diversity and admixture were analysed using ANGSD[62] (Supplementary note).

Mitochondrial DNA data processing

The mitochondrial genome exhibited an average coverage depth of 322× (ranging from 24x to 5,868x) per sample (Supplementary Table 1). Mitochondrial reads were extracted and filtered from poorly mapped reads using samtools view (-q 20 -f 0x0002 -F 0x0004 -F 0x0008) and from duplicated reads using Picard v.2_14_0 MarkDuplicates (https://github.com/broadinstitute/picard). Realignment around indels was applied with GATK[63] and SNP were subsequently called using samtools[64] mpileup using only reads that acheived a mapping quality of 30 and base quality of 30. The occurrence of heterozygous sites in mtDNA, known as heteroplasmy, has been described across vertebrate species[65] and in other nematode species, including *C. briggsae*[66] and *C. elegans*[67]. However, heterozygous sites may also occur as technical artefacts as a result of genetically similar sequences shared between the nuclear and mitochondrial genomes, *i.e.*, numts[68]. To exclude sites prone to heterozygous signals from further phylogenetic inference analysis, a SNP calling procedure was implemented with the HaplotypeCaller tool to apply hard filtering parameters on the raw SNP sets (QD>=10, FS<=35 MQ>=30 MQRankSum>=-12.5 and ReadPosRankSum >=-8) with a minimum depth of 20 reads. This procedure excluded 1,354 putative heterozygous sites, and retained 72%

of the putative SNP sites (3,052 out of 4,234 SNPs). Nucleotide diversity and Tajima's *D* were computed by sliding-windows of 100 bp using vcftools v.0.1.15[69]. A principal component analysis (PCA) was performed on genotypes using the SNPrelate package[70] in R version 3.5[71]. A consensus fasta sequence was subsequently generated with GATK FastaAlternateReferenceMaker for each sample using the filtered variant set, which was used for the phylogenetic analyses.

Diversity and divergence analysis

Genome-wide nucleotide diversity (π) was computed for each population with at least five individuals using ANGSD[62]. Using genotype likelihoods (GLs) from samtools[64] (option GL=1) as an input, variants were included that had a minimal supporting evidence of 5 reads, and base and mapping quality phred scores of at least 20. As π values were biased by population mean coverage (Supplementary Fig. 17), π was also calculated from a subset of populations containing individuals with a minimum mean coverage of 5×: this was limited to France (FRA.1, n = 5, mean coverage of 7.66×), Guadeloupe (n = 5, mean coverage of 12.75×) and Namibia (n = 6, mean coverage of 9.85×).

F_{ST} was estimated from the VQSR-called genotypes between populations with at least five individuals using the Weir-Cockerham estimator[72] in vcftools v0.1.15[69]. To prevent artifactual signal linked to variation in coverage between populations, F_{ST} was calculated on subsets of SNPs, binned based on their minor allele frequency (MAF) in 10% increments. The maximum F_{ST} value calculated was retained for comparison (Supplementary Fig. 4) and the maximal value was considered as reported elsewhere[73]. The resulting F_{ST} estimates were not biased by coverage, as measured by negligible correlation between pairwise F_{ST} coefficients and associated

population cross-coverage (Pearson's $r_{(136)}$ = -0.05, P = 0.55; Supplementary Fig. 18).

The pairwise sequence divergence between individual samples was calculated using the Hamming distance, i.e. 1-IBS, using PLINK[74], considering the only individuals with a mean coverage above 2.5× as failure to do so yielded biased estimates (Supplementary note, Supplementary Fig. 19). A neighbour-joining tree of Hamming distances calculated from the nuclear DNA genotypes was built using the R package *ape*[75].

PCA on genotypes inferred from genotype likelihoods of the 223 samples was generated using ANGSD ngsCovar, filtering for sites with base and mapping quality phred scores less than 30, minimum depth of 5 reads, and a SNP *p*-value (as computed by ANGSD) below 10^{-3}. Clustering was robust to coverage variation and closely matched the PCA from the VQSR-called SNP genotypes (Supplementary Fig. 20).

Phylogenetic inference

To determine the phylogenetic structure of the cohort, the 223 consensus mitochondrial fasta sequences, including the mtDNA genomes of *Haemonchus placei* (accession number AP017687.1) and *Teladorsagia circumcincta* (accession number AP017699.1) as outgroup sequences, were first aligned using Muscle v3.8.31[76], followed by stringent trimming of sequence alignments using Gblocks[77]. The most likely evolutionary model, GTR substitution model with rate heterogeneity modelled by a gamma distribution with invariable sites, was determined using modelgenerator v.0.851[78]. A maximum-likelihood tree was subsequently generated

using PhyML[79] v.20120412, with branch supports computed using 100 bootstraps.

Admixture analysis

Admixture was determined using NGSAdmix[80]. This tool relies on genotype likelihoods to account for data uncertainty, and has been shown to produce robust inferences about population ancestry from low coverage samples alone, or a mixture of low and higher coverage samples[80]. This analysis was performed on 223 samples, for *K* ranging from 2 to 10 clusters (Supplementary Fig. 6 and 7), retaining sites with less than 50% missing data across individuals and minor allele frequency (MAF) above 5%. Five iterations were run leaving one autosome out at a time, and the best K was chosen as the first value that would minimize the median absolute deviation across runs (Supplementary Fig. 6). Sample coverage did not affect the results (Supplementary Fig. 21).

Effective population size and population divergence dating

The effective population size (*Ne*) trajectory through time, and the cross-divergence time between populations, were estimated using MSMC[81]. This approach uses patterns of heterozygosity along the genome to identify past recombination events modelled as Markov processes[82]. Mutation density along the sequence mirrors either recent (long tract of limited diversity) or older (enrichment in heterozygosity over short distances) events. According to coalescent theory[83], at any given time, the amount of recombination is proportional to *Ne*.

MSMC was applied to individuals with a mean coverage above 10×, limiting the analysis to six populations (FRA.1, FRG, IND, MOR, NAM, STA.1), and considering four haplotypes per population. Beagle v4.1 was used to impute missing genotypes

and to establish phase in VQSR SNP calls. Imputation accuracy analysis revealed a 7.2% and 9.0% discordance rates at the individual and site levels respectively (Supplementary Fig. 22). For each population, estimates were averaged across five runs, leaving one chromosome out at a time for cross-validation. MSMC times and coalescent rates were scaled to real time and population sizes by considering a mutation rate similar to that of *C. elegans*, *i.e.* 2.7×10^{-9} per generation[18], a balanced sex-ratio[19] and an inferred generation interval of 40 days (the sum of 10 days to reach mature free-living infective larvae from the egg stage, and a 30-day prepatent period for fully mature egg-laying females)[84].

Additional support to the estimates from the nuclear genome were obtained from a phylogenetic analysis of coding sequences using BEASTv1.10[85]. Mitochondrial coding sequences were extracted from the consensus sequence of every individual, concatenated per individual, and aligned using Musclev3.8.31[76]. A Bayesian skyline model[86] was used, with a HKY substitution model and a strict clock model, as other modalities yielded weak effective sample size (ESS) and unstable parameter values. Clock rate was set to *C. elegans* mitochondrial mutation rate, i.e. 1×10^{-7} per site per generation[67], as variation in sampling date was not sufficient to estimate molecular rate. Parameters showed sufficient sampling (effective size above 200) after 50,000,000 iterations and a burn-in of the first 20 million steps. Node ages were scaled to years assuming a generation interval of 40 days. A maximum clade credibility tree was generated with TreeAnnotator v1.10.1 (http://beast.community/treeannotator).

Diversifying selection scan

To further characterize the genetic diversity in *H. contortus* populations, we identified

genomic regions under diversifying selection using XP-CLR[26]. This approach takes advantage of both within-population distortion of the allele frequency spectrum, and between-population differences in allele frequencies in the vicinity of selective sweeps[26]. Although XP-CLR is robust to SNP ascertainment bias[26], the analysis was restricted to populations with at least five individuals with a minimum mean coverage of 2× per individual. Unphased VQSR-derived genotypes were filtered to retain SNPs with a within-population call rate >80% and MAF >5%. The analysis was run on every one of the 22 possible within-continent pairwise comparisons of the retained populations (Australia and Africa) with the following options: -w1 0.0001 500 2000 –p0 0. This fit a grid of putatively selected points every 2 Kbp along the genome, with a sliding window size around grid points of 0.01 cM, interpolating SNP position in Morgans from the average recombination rate for each chromosome[12]. Down-sampling was applied to windows where more than 500 SNPs were found to keep SNP numbers comparable between regions, and no LD-based down-weighing of the CLR scores was applied. A selection score was subsequently computed at every position as the root mean square of XP-CLR coefficients, and the highest 0.1% of selection scores were deemed significant, as reported elsewhere[73].

Analysis of β-tubulin isotype 1 (Hco-btub-iso-1) and the genetic architecture of benzimidazole- and ivermectin-resistance

The genomic coordinates of *Hco-btub-iso-1* were determined by blasting the gene coding sequence from WormBase Parasite[87] against current genome assembly (BLASTN, e-value < 10^{-50}). The SNP positions associated with codons 167, 198 and 200 was determined to be at 7027535, 7027755, and 7027758 bp, respectively, along chromosome I after aligning the *Hco-btub-iso-1* consensus sequence and published sequence[88] (GenBank accession FJ981629.1) using muscle v3.8.31[76].

Genotype and GLs at these positions were determined with ANGSD[62]. Genotypes were only considered for GL >60% (n = 74) as coverage bias occurred otherwise (Supplementary Table 5 and Supplementary Fig. 11). For these samples, phased and imputed genotypes from VQSR SNPs spanning the *Hco-btub-iso-1* locus were used to compute pairwise number of allele differences.

The introgression of resistant haplotypes from France mainland into Guadeloupian worm populations was tested by a topology weighting analysis implemented with TWISST[89]. This method computes phylogenetic trees between a set of populations, using genetic information from short sliding windows (50 Kbp) and by sampling with replacement individuals from each population. At each window, a weight is subsequently computed for each of the possible tree topology, ultimately providing inference of introgression events in discrete locations where tree topology connects phylogenetically distant populations. Analyses were run using individuals with sufficient mean depth of coverage (5× and more) from French (FRA.1; n = 5; mean depth = 7.62×) and Guadeloupian (FRG; n = 5; mean depth = 12.75×) populations, adding Namibian population (n = 2; mean depth = 10.9×) as an outgroup. A fourth population was chosen for its common ancestry with Guadeloupian populations: first, an analysis was run with worms from São Tomé (n = 2; mean depth = 7.01×), followed by a second analysis using Moroccan samples (n = 2; mean depth = 10.6×), to ensure that evidence of introgression was consistent in both cases.

To investigate the genetic architecture of ivermectin resistance, a differentiation scan was run between ivermectin-resistant populations and populations of unknown status sharing same ancestry, i.e. STA.1 and STA.3 versus NAM and ZAI in Africa, AUS.1

against AUS.2 in Australia. Windowed F_{ST} estimates[72] were calculated every 10 Kbp with a 1 Kbp overlap along the genome with ANGSD[62], retaining sites with minimal depth of 5 reads, mapping and base quality phred scores of at least 30, missing rate below 50%, and windows with at least 1000 sites. Genomic coordinates of the top 0.5% most differentiated windows were extracted and analysed by BLASTN (minimum e-value = 10^{-50}) against the published V1 *H. contortus* assembly[13], and annotated gene identifiers were inferred from the corresponding GFF file from WormBase Parasite[87].

Identification of SNPs under environmental selection

To identify SNPs putatively influenced by environmental selection, we first defined the climatic conditions of each population using the Köppen-Geiger classification[90], and inferred from population geographical coordinates[91]. Three populations with the best coverage, i.e. Namibia, France, and Guadeloupe, were used to contrast dry, temperate and tropical conditions, respectively. We performed a genome-wide scan using pairwise comparisons of F_{ST} using ANGSD[62], of which test values greater than the 0.5% quantile in at least two comparisons were analysed further.

To further explore the contribution of environmental climatic variables on standing genetic variation, we applied a machine-learning gradient forest algorithm[92] to quantify changes in genetic variation (fit as population SNP MAF) along environmental gradients. The gradients consisted of 19 bioclimatic variables[93] summarizing rainfall and temperature information (supplementary table 8). For each SNP, a random forest of 500 trees was grown. For each tree, bootstrapped SNP MAFs were regressed against a random subset of bioclimatic variables to determine the variable that best partitioned the data, thereby building the first node that

partitions the data into two sets of homogeneous observations. Iterations follow to determine subsequent nodes by resampling a random subset of bioclimatic variables until no observations are left. The proportion of variance explained by each bioclimatic variable is then averaged across SNPs, and the function of SNP frequency modification along bioclimatic variable is built.

The analysis was performed populations with at least 5 individuals, a mean coverage of 2×, and no record of ivermectin resistance (AUS.2, FRA.1, GUA, IND, MOR, NAM, STO, ZAI). SNP MAFs were estimated from genotype likelihoods with ANGSD, and subsequently filtered to ensure a within-population MAF >10%, at least 90% within-population call-rate, and shared across the eight considered populations, resulting in 3,758 SNPs retained for further analysis.

Environmental variables were highly correlated (Supplementary Fig. 14a), resulting in instability in the predictor's importance. To minimise this effect, the gradient forest analysis was restricted to the 11 environmental variables least correlated as assessed by a PCA (Supplementary Fig. 14b). SNP-environment associations were further investigated by focusing on the top two environmental predictors of genomic variations and using a Latent Factor Mixed Model analysis. This analysis was implemented with the lfmm R package[94] on genotype calls from the VQSR pipeline, considering SNPs with call rate of at least 70% across populations and minor genotype frequency above 5%, using K = 3 for the latent factor accounting for underlying population structure. Analyses were run 5 times and *P* values were combined and adjusted as recommended[94].

Gene identification and Gene Ontology enrichment analysis

Annotation of the revised *H. contortus* genome is ongoing. Therefore, genes underlying major differentiation signals were inferred by blasting the region of interest (10 Kbp window for F_{ST} analyses, 2000 bp region around XP-CLR hits) against the published V1 *H. contortus* genome assembly[13]. Any genes falling within most probable blast hit coordinates were retrieved from the *H. contortus* published assembly available from WormBase Parasite[87]. This database was also used to retrieve *C. elegans* orthologs of positional candidate genes and *H. contortus* Gene Ontology (GO) terms. GO term enrichment analysis was run with the R topGO[95] package, considering nodes with at least 10 annotated genes. The “weight01” algorithm was used to account for existing topology between GO terms. This framework makes *P*-value of one GO term conditioned on its neighbours, thereby correcting for multiple testing. Enrichment was tested by the Kolmogorov-Smirnov statistic applied to gene selection or F_{ST} score accordingly. GO terms with *P* values below 1% were deemed significant.

References

1. Blaxter, M. & Koutsovoulos, G. The evolution of parasitism in Nematoda. *Parasitology* **142 Suppl 1**, S26-39 (2015).
2. O'Connor, L.J., Walkden-Brown, S.W. & Kahn, L.P. Ecology of the free-living stages of major trichostrongylid parasites of sheep. *Vet Parasitol* **142**, 1-15 (2006).
3. Jones, J.T. *et al.* Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol Plant Pathol* **14**, 946-61 (2013).
4. Anderson, R.C. The origins of zooparasitic nematodes. *Can J Zool* **62**, 317-328 (1984).
5. Kaplan, R.M. & Vidyashankar, A.N. An inconvenient truth: Global worming and anthelmintic resistance. *Veterinary Parasitology* **186**, 70-78 (2012).
6. McKellar, Q.A. & Jackson, F. Veterinary anthelmintics: old and new. *Trends Parasitol* **20**, 456-61 (2004).
7. Bundy, D.A.P. *et al.* 100 Years of Mass Deworming Programmes: A Policy Perspective From the World Bank's Disease Control Priorities Analyses. *Adv Parasitol* **100**, 127-154 (2018).
8. Schulz, J.D., Moser, W., Hurlimann, E. & Keiser, J. Preventive Chemotherapy in the Fight against Soil-Transmitted Helminthiasis: Achievements and Limitations. *Trends Parasitol* (2018).
9. Hewitson, J.P. & Maizels, R.M. Vaccination against helminth parasite infections. *Expert Rev Vaccines* **13**, 473-87 (2014).
10. Sallé, G. *et al.* Transcriptomic profiling of nematode parasites surviving vaccine exposure. *Int J Parasitol* (2018).
11. Gilleard, J.S. Haemonchus contortus as a paradigm and model to study anthelmintic drug resistance. *Parasitology* **140**, 1506-22 (2013).

12. Doyle, S.R. *et al.* 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 (2018).

13. Laing, R. *et al.* The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. *Genome Biol* **14**, R88 (2013).

14. International Helminth Genomes Consortium, Coghlan, A., Mitreva, M. & Berriman, M. Comparative genomics of the major parasitic worms. *bioRxiv* (2018).

15. Mougin, C. *et al.* BRC4Env, a network of Biological Resource Centres for research in environmental and agricultural sciences. *Environ Sci Pollut Res Int* (2018).

16. Langley, C.H. *et al.* Genomic variation in natural populations of Drosophila melanogaster. *Genetics* **192**, 533-98 (2012).

17. Perry, G.H. *et al.* Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res* **22**, 602-10 (2012).

18. Denver, D.R. *et al.* A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. *Proc Natl Acad Sci U S A* **106**, 16310-4 (2009).

19. Saccareau, M. *et al.* Meta-analysis of the parasitic phase traits of Haemonchus contortus infection in sheep. *Parasit Vectors* **10**, 201 (2017).

20. Geggus, D. The French Slave Trade: An Overview. *The William and Mary Quaterly* **58**, 119-138 (2001).

21. Anonymous. The wool industry - looking back and forward. (Australian Bureau of Statistics, 2003).

22. Evesson, B. & Moor, R. The foundation of Australia's fine wool industry. in *Journal of the Royal Australian Historical Society* (2000).

23. Doyle, S.R. *et al.* A major locus for ivermectin resistance in a parasitic nematode. *bioRxiv* (2018).

24. Eng, J.K. *et al.* Ivermectin selection on beta-tubulin: evidence in Onchocerca volvulus and Haemonchus contortus. *Mol Biochem Parasitol* **150**, 229-35 (2006).

25. de Lourdes Mottier, M. & Prichard, R.K. Genetic analysis of a relationship between macrocyclic lactone and benzimidazole anthelmintic selection on Haemonchus contortus. *Pharmacogenet Genomics* **18**, 129-40 (2008).

26. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res* **20**, 393-402 (2010).

27. Ardelli, B.F., Stitt, L.E., Tompkins, J.B. & Prichard, R.K. A comparison of the effects of ivermectin and moxidectin on the nematode Caenorhabditis elegans. *Vet Parasitol* **165**, 96-108 (2009).

28. Janssen, I.J., Krucken, J., Demeler, J. & von Samson-Himmelstjerna, G. Transgenically expressed Parascaris P-glycoprotein-11 can modulate ivermectin susceptibility in Caenorhabditis elegans. *Int J Parasitol Drugs Drug Resist* **5**, 44-7 (2015).

29. Avery, L. & Horvitz, H.R. Effects of starvation and neuroactive drugs on feeding in Caenorhabditis elegans. *J Exp Zool* **253**, 263-70 (1990).

30. Scott, E.W., Baxter, P. & Armour, J. Fecundity of anthelmintic resistant adult Haemonchus contortus after exposure to ivermectin or benzimidazoles in vivo. *Res Vet Sci* **50**, 247-9 (1991).

31. Kotze, A.C. & Prichard, R.K. Anthelmintic Resistance in Haemonchus contortus: History, Mechanisms and Diagnosis. *Adv Parasitol* **93**, 397-428 (2016).

32. Martin, R.J. *et al.* Levamisole receptors: a second awakening. *Trends Parasitol* **28**, 289-96 (2012).

33. Wada, M. *et al.* Isolation and characterization of a GDP/GTP exchange protein specific for the Rab3 subfamily small G proteins. *J Biol Chem* **272**, 3875-8 (1997).

34. Cao, R. & Zhang, Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**, 155-64 (2004).

35. Blouin, M.S., Yowell, C.A., Courtney, C.H. & Dame, J.B. Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* **141**, 1007-14 (1995).

36. Dey, A., Chan, C.K., Thomas, C.G. & Cutter, A.D. Molecular hyperdiversity defines populations of the nematode Caenorhabditis brenneri. *Proc Natl Acad Sci U S A* **110**, 11056-60 (2013).

37. Gilleard, J.S. & Redman, E. Genetic Diversity and Population Structure of Haemonchus contortus. *Adv Parasitol* **93**, 31-68 (2016).

38. Larson, G. *et al.* Current perspectives and the future of domestication studies. *Proc Natl Acad Sci U S A* **111**, 6139-46 (2014).

39. Pleurdeau, D. *et al.* "Of sheep and men": earliest direct evidence of caprine domestication in southern Africa at Leopard Cave (Erongo, Namibia). *PLoS One* **7**, e40340 (2012).

40. Campbell, M.C. & Tishkoff, S.A. The evolution of human genetic and phenotypic variation in Africa. *Curr Biol* **20**, R166-73 (2010).

41. Chessa, B. *et al.* Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**, 532-6 (2009).

42. Tapio, M. *et al.* Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Mol Biol Evol* **23**, 1776-83 (2006).

43. Kijas, J.W. *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol* **10**, e1001258 (2012).

44. Spangler, G.L. *et al.* Whole genome structural analysis of Caribbean hair sheep reveals quantitative link to West African ancestry. *PLoS One* **12**, e0179021 (2017).

45. Naves, M., Alexandre, G., Leimbacher, F., Mandonnet, N. & Menendez-buxadera, A. Les ruminants domestiques de la Caraïbe: le point sur les ressources génétiques et leur exploitation. *INRA Prod. Anim.* **14**, 181-192 (2001).

46. Roman, A. *Saint Malo au temps des négriers*, 357 (Karthala, 2003).

47. Pereira, F. *et al.* Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Mol Biol Evol* **23**, 1420-6 (2006).

48. Redman, E. *et al.* The emergence of resistance to the benzimidazole anthlemintics in parasitic nematodes of livestock is characterised by multiple independent hard and soft selective sweeps. *PLoS Negl Trop Dis* **9**, e0003494 (2015).

49. Vercruysse, J. *et al.* Is anthelmintic resistance a concern for the control of human soil-transmitted helminths? *Int J Parasitol Drugs Drug Resist* **1**, 14-27 (2011).

50. Laing, R., Gillan, V. & Devaney, E. Ivermectin - Old Drug, New Tricks? *Trends Parasitol* **33**, 463-472 (2017).

51. Osei-Atweneboana, M.Y. *et al.* Phenotypic evidence of emerging ivermectin resistance in Onchocerca volvulus. *PLoS Negl Trop Dis* **5**, e998 (2011).

52. Doyle, S.R. *et al.* Genome-wide analysis of ivermectin response by Onchocerca volvulus reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. *PLoS Negl Trop Dis* **11**, e0005816 (2017).

53. Rajpurohit, S., Oliveira, C.C., Etges, W.J. & Gibbs, A.G. Functional genomic and phenotypic responses to desiccation in natural populations of a desert drosophilid. *Mol Ecol* **22**, 2698-715 (2013).

54. Sharma, V., Kohli, S. & Brahmachari, V. Correlation between desiccation stress response and epigenetic modifications of genes in Drosophila melanogaster: An example of environment-epigenome interaction. *Biochim Biophys Acta* **1860**, 1058-1068 (2017).

55. Coustham, V. *et al.* Quantitative modulation of polycomb silencing underlies natural variation in vernalization. *Science* **337**, 584-7 (2012).

56. Cassada, R.C. & Russell, R.L. The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans. *Dev Biol* **46**, 326-42 (1975).

57. Blitz, N.M. & Gibbs, H.C. Studies on the arrested development of Haemonchus contortus in sheep. I. The induction of arrested development. *Int J Parasitol* **2**, 5-12 (1972).

58. Wood, D.E. & Salzberg, S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**, R46 (2014).

59. Li, B.W. *et al.* Transcriptomes and pathways associated with infectivity, survival and immunogenicity in Brugia malayi L3. *BMC Genomics* **10**, 267 (2009).

60. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).

61. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint* **arXiv:1207.3907 [q-bio.GN]**(2012).

62. Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

63. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).

64. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).

65. Rensch, T., Villar, D., Horvath, J., Odom, D.T. & Flicek, P. Mitochondrial heteroplasmy in vertebrates using ChIP-sequencing data. *Genome Biol* **17**, 139 (2016).

66. Howe, D.K., Baer, C.F. & Denver, D.R. High rate of large deletions in Caenorhabditis briggsae mitochondrial genome mutation processes. *Genome Biol Evol* **2**, 29-38 (2009).

67. Konrad, A. *et al.* Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in Caenorhabditis elegans Spontaneous Mutation Accumulation Lines of Differing Population Size. *Mol Biol Evol* **34**, 1319-1334 (2017).

68. Hazkani-Covo, E., Zeller, R.M. & Martin, W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* **6**, e1000834 (2010).

69. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).

70. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-8 (2012).

71. R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, 2016).

72. Weir, B.S. & Cockerham, C.C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370 (1984).

73. Svardal, H. *et al.* Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. *Nat Genet* **49**, 1705-1713 (2017).

74. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

75. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289-90 (2004).

76. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004).

77. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-77 (2007).

78. Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J. & McLnerney, J.O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* **6**, 29 (2006).

79. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).

80. Skotte, L., Korneliussen, T.S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693-702 (2013).

81. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**, 919-25 (2014).

82. McVean, G.A. & Cardin, N.J. Approximating the coalescent with recombination. *Philos Trans R Soc Lond B Biol Sci* **360**, 1387-93 (2005).

83. Kingman, J.F.C. The coalescent. *Stochastic Processes and their Applications* **13**, 235-248 (1982).

84. Emery, D.L., Hunt, P.W. & Le Jambre, L.F. Haemonchus contortus: the then and now, and where to from here? *Int J Parasitol* **46**, 755-769 (2016).

85. Suchard, M.A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016 (2018).

86. Drummond, A.J., Rambaut, A., Shapiro, B. & Pybus, O.G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185-92 (2005).

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

88. Rufener, L., Kaminsky, R. & Maser, P. In vitro selection of Haemonchus contortus for benzimidazole resistance reveals a mutation at amino acid 198 of beta-tubulin. *Mol Biochem Parasitol* **168**, 120-2 (2009).

89. Martin, S.H. & Van Belleghem, S.M. Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* **206**, 429-438 (2017).

90. Köppen, W. Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet *Meteorol. Z.* **1**, 215-226 (1884).

91. Chen, D. & Chen, H.W. Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development* **6**, 69-79 (2013).

92. Ellis, N., Smith, S.J. & Pitcher, C.R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156-68 (2012).

93. Fick, S., E. & Hijmans, R.J. WorldClim 2: new 1□km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302-4315 (2017).

94. Frichot, E., Schoville, S.D., Bouchard, G. & Francois, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* **30**, 1687-99 (2013).

95. Alexa, A., Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. (2016).

Acknowledgements

JAC, MB, NH, and SD are supported by the Wellcome Trust via their core funding of the Wellcome Trust Sanger Institute (grant 206194). GS has received the support of the EU in the framework of the Marie-Curie FP7 COFUND People Programme, through the award of an AgreenSkills (grant agreement n° 267196) and AgreenSkills+ fellowships (grant agreement n°609398).

Author contributions

GS, JAC designed the experiment. GS, JAC, SD drafted the manuscript. JCa and JCo sampled and prepared parasite materials. GS performed DNA extraction and data analyses. SD built the reference genome. MB and NH managed and supervised parasite sequencing and the project.

Competing interests

Authors declare they have no competing interests.

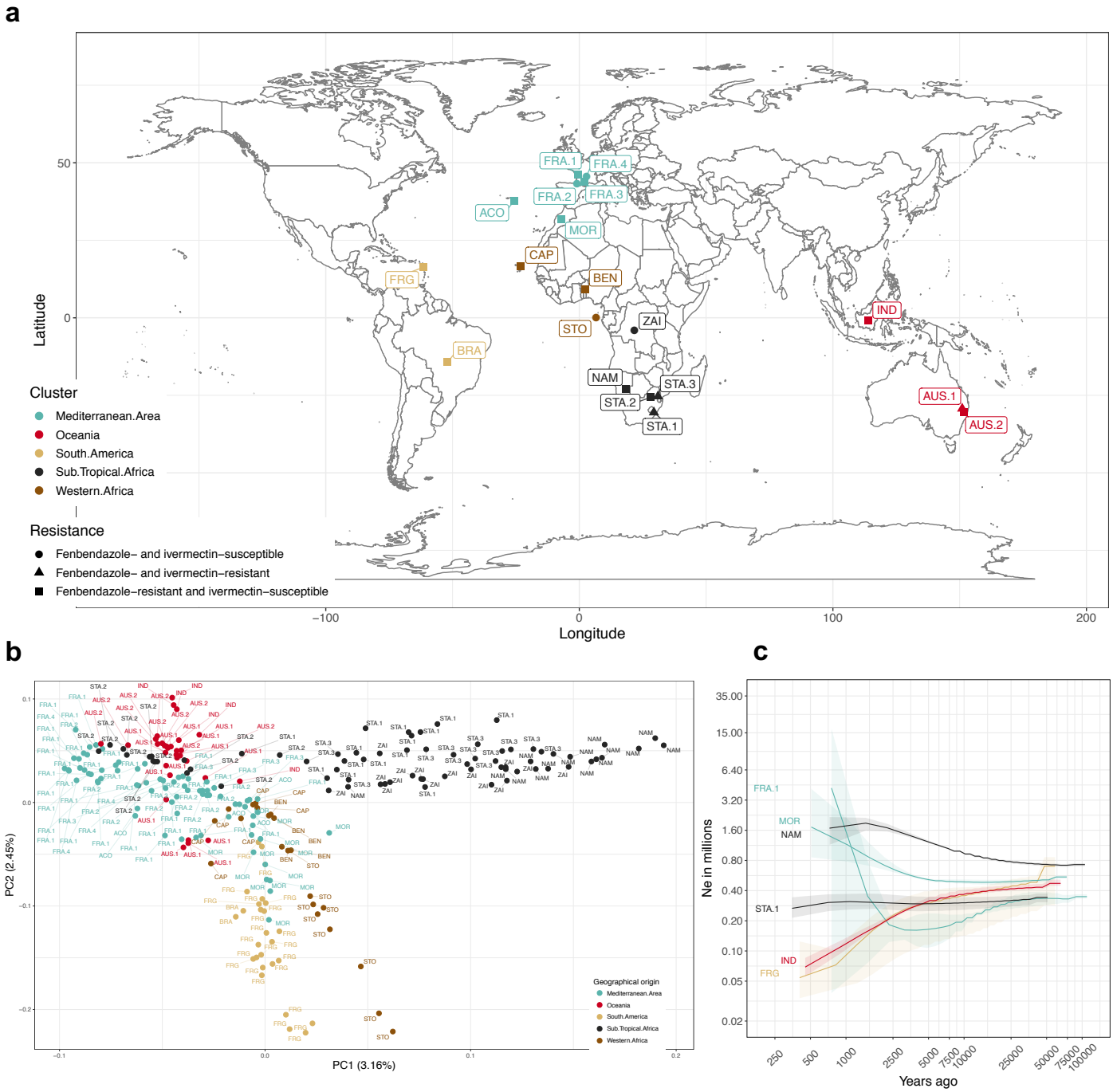


Figure 1. Global diversity of *Haemonchus contortus.*

(**a**) Map depicts the global distribution of the populations sampled, coloured by geographical region (sand: South-America, brown: Western-Africa, dark green: Mediterranean area, black: Subtropical Africa, red: Australia). Shape indicates population resistance status to anthelmintics, either resistant (square) or susceptible (circle) to fenbendazole or resistant to ivermectin and fenbendazole (triangle).

(**b**) Principal component analysis based on genotype likelihood inferred from whole genome sequences of 223 individual males (1,638,715 variants considered). Samples are coloured by geographic region described in panel 1a.

(**c**) MSMC effective population size across time for 6 populations. Coloured shaded area represents range of values estimated from a cross-validation procedure with five replicates, computed by leaving one chromosome out at a time.

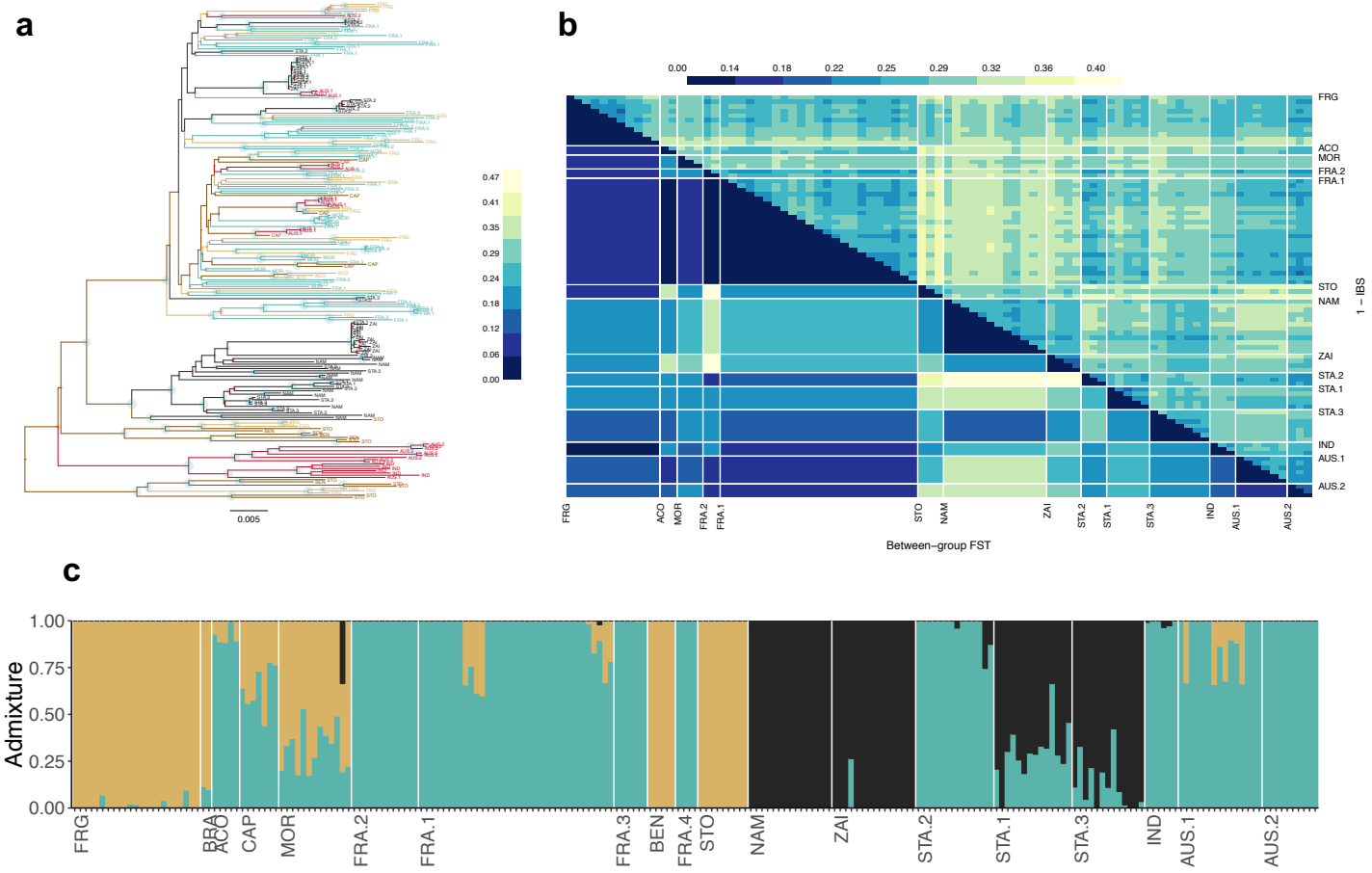


Figure 2. Global connectivity of *Haemonchus contortus* populations.
(a) Maximum likelihood phylogenetic tree based on 223 individual mitochondrial genomes, rooted on *Teladorsagia circumcincta* and *Haemonchus placei*. Circles indicate bootstrap support for each branch, blue if support was higher than 50%, red elsewhere. The root node has a 84% support. Branches leading to a sample identifier are coloured by geographic region described in Figure 1a.
(**b**) Matrix showing pairwise dissimilarity between individuals (upper right) and pairwise F_{ST} between populations (lower left).
(**c**) Admixture analysis for K = 3 clusters and for 223 individuals. Populations are presented sorted along their longitudinal range, and samples sorted by assignment to the 3 clusters.

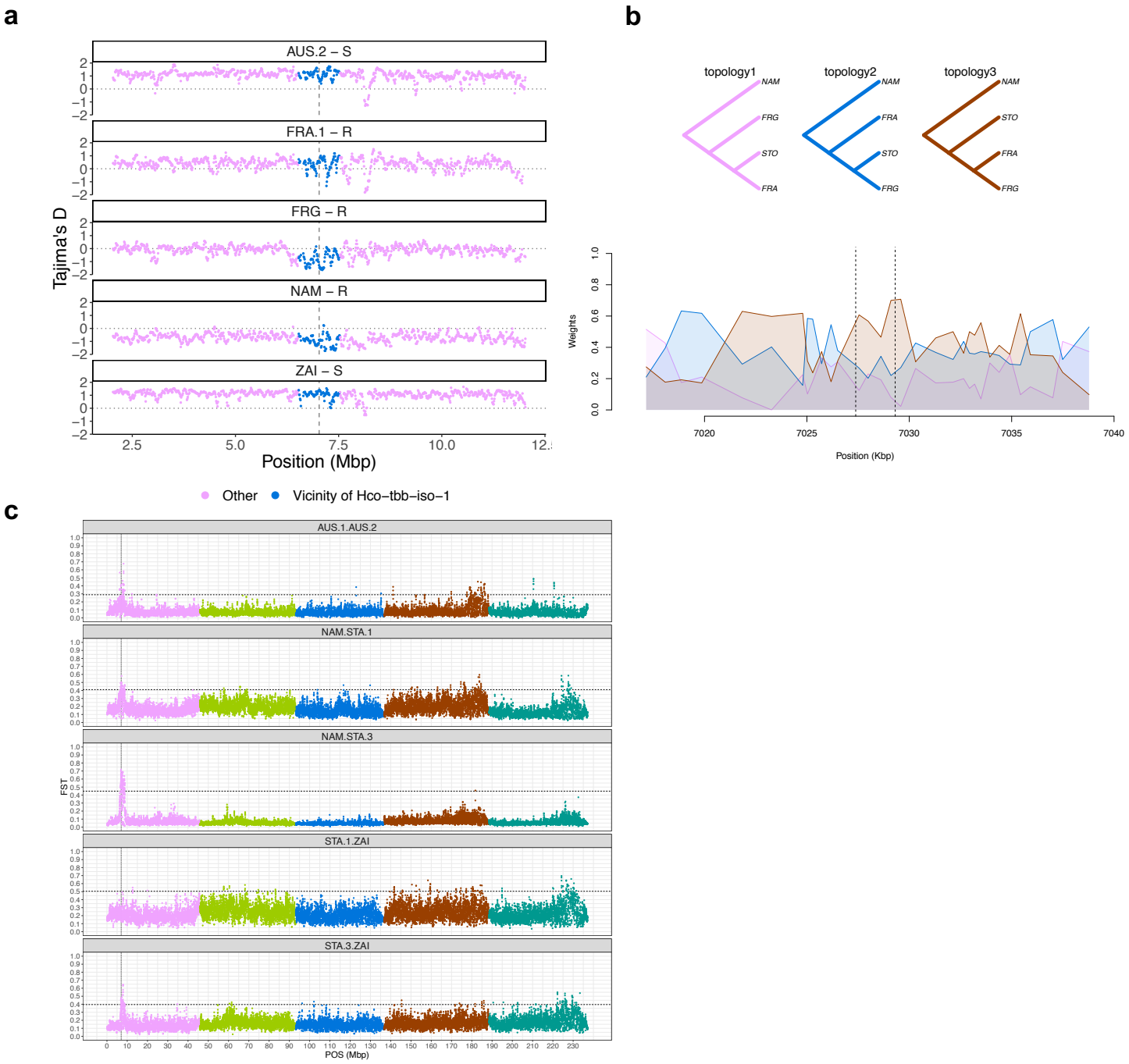
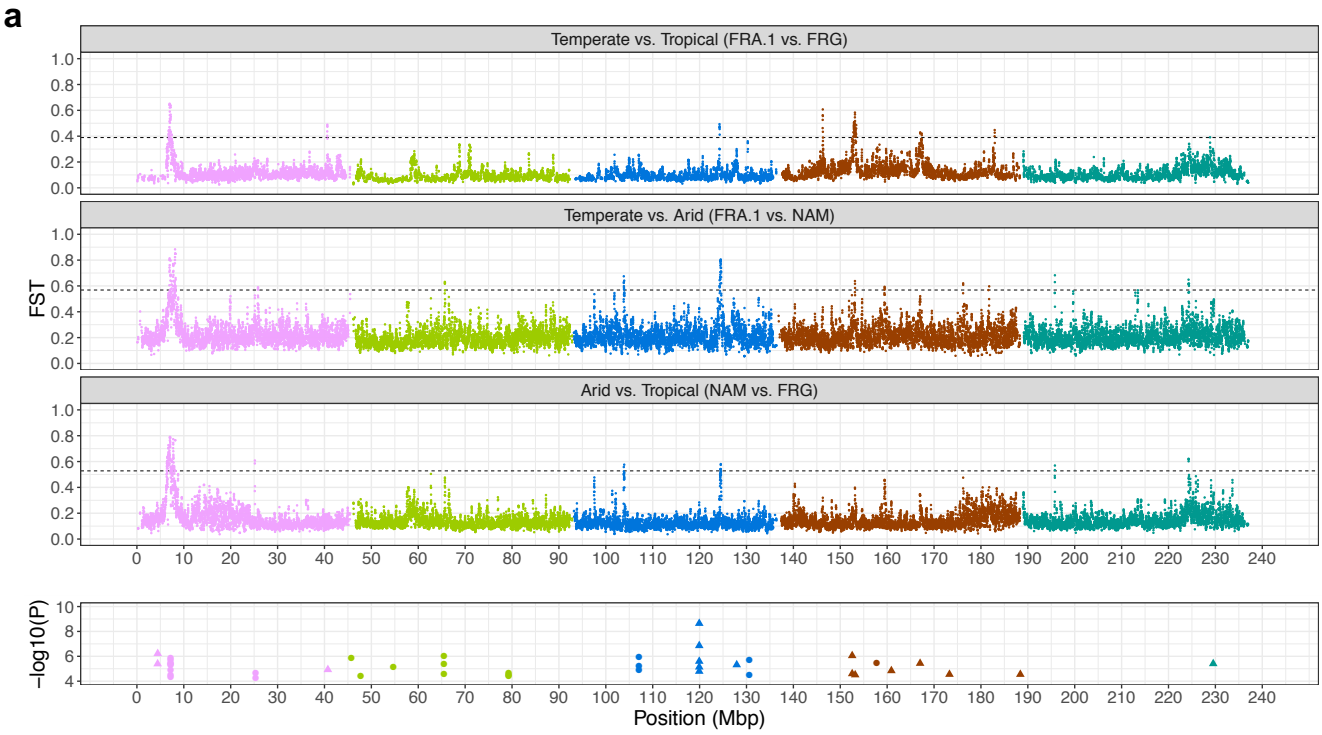


Figure 3. Anthelmintic-mediated selection is a major driver of genetic variation

(**a**) Windowed Tajima's D estimated across a 10-Mbp window of chromosome I (pink) and within 1 Mbp surrounding *Hco-tbb-iso-1* (blue) for French, Guadeloupian and Namibian populations.

(**b**) Topology weighing analysis of a 20-Kbp window centred on *Hco-tbb-iso-1*. At each position, the weight of each of the three possible topologies inferred from 50 Kbp-windows are overlaid. Topology 2 (blue) corresponds to an isolation-by-distance history, while topology 3 (brown) would agree with shared introgressed material between worm populations from French mainland into Guadeloupe. The *Hco-tbb-iso-1* locus is indicated by vertical dashed lines.

(**c**) Genome-wide differentiation scan between pairs of ivermectin-resistant (AUS.1, STA.1, STA.3) and populations of unknown status. Chromosomes are coloured and ordered by name, i.e. from I (pink) to V (forest green). Horizontal line represents the 0.5% F_{ST} quantile cut-off. Vertical line on chromosome I points at *Hco-tbb-iso-1* locus.



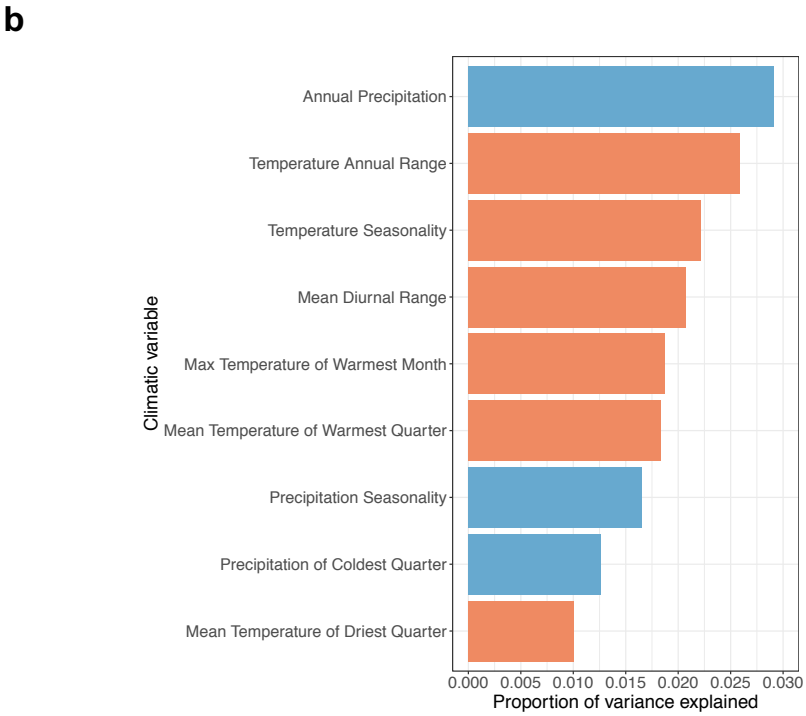


Figure 4. Genomic signal of climate adaptation

(**a**) Top three panels represent F_{ST} values plotted against genomic position (Mbp), colored and ordered by chromosome name (from I to V). Horizontal dashed line represents the 0.5% quantile. Bottom panel shows genomic positions of significant associations between annual precipitation (circle) and temperature annual range (triangle).

(**b**) Proportion of genetic variance (within 8 *H. contortus* populations retained for analysis), explained by precipitation- (blue) and temperature-associated (orange) climatic variables. Analysis was run on a set of nine variables out of 19, selected to minimize collinearity issue.)