The global diversity of the major parasitic nematode *Haemonchus contortus* is shaped by human intervention and climate

Sallé G.[1,2,*], Doyle S.R.[1], Cortet J.[2], Cabaret J., 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 an haematophagous parasitic nematode of veterinary interest and a model for the study of drug resistance mechanisms or host-parasite interactions. To understand its evolutionary history, and its ability to adapt in the face of climatic and drug pressure, we have performed an extensive survey of genome-wide diversity using single-worm whole genome sequencing of 223 individuals sampled from 19 isolates spanning five continents. The pattern of global diversity is driven by an African origin for the species, together with contemporary dispersal that is consistent with modern human movement, with

evidence for parasites spreading during the transatlantic slave trade and colonisation of Australia presented. Strong selective sweeps were identified in independent populations each surrounding the β-tubulin locus, a target of benzimidazole anthelmintic drug treatment used widely to control *H. contortus* infections. These signatures of selection were further supported by signals of diversifying selection 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 identify some known, and previously undescribed, candidate genes that may play a role in ivermectin resistance. Finally, we describe genetic signatures of climate-driven adaptation, revealing a gene acting as an epigenetic regulator and components of the *dauer* pathway may play a role in adaptation in the face of climatic fluctuations. These results begin to define genetic adaptation to climate for the first time 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. Although 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]. Their complex life cycles 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 have adapted to a wide range of threats, including predation, climate and the immune responses of a great diversity of both plant and animal host species[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 animal parasitic nematodes 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 veterinary parasites [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

66 licensed vaccines are currently available for veterinary purposes[10], transcriptomic
67 plasticity of the parasite following vaccine challenge may contribute to circumvent the
68 vaccinal response of their host[10]. It is therefore clear that novel, sustainable control
69 strategies are required. The potential of helminths to adapt to – and thus escape –
70 control measures lies in their underlying genetic diversity. A greater understanding of
71 the extent of this diversity and the processes that shape it throughout their range
72 should provide insight into the mechanisms by which they adapt, and may identify
73 new targets which may be exploited for control.
74
75 The trichostrongylid *Haemonchus contortus* is a gastrointestinal parasite of
76 ruminants in tropical and temperate regions throughout the world, and causes
77 significant economic and animal health burden particularly on sheep husbandry. It is
78 also emerging as a model parasitic nematode system for functional and comparative
79 genomics, largely due to its rapid ability to acquire drug resistance, the relative
80 tractability of its life-cycle under laboratory conditions[11], the development of
81 extensive genomic resources[12,13], and its relatively close relationship with other
82 clade V parasitic nematodes of both veterinary and medical importance (i.e, other
83 gastro-intestinal nematodes of livestock and human hookworms)[14]. We have used
84 whole genome sequencing of 223 individual *H. contortus* sampled from 19 isolates
85 spanning five continents to characterise genome- and population-wide genetic
86 diversity throughout its range. This survey of genome-wide diversity has revealed old
87 and new genetic connectivity influenced by human history, and signatures of
88 selection in response to anthelmintic exposure and local climatic variation.

Results

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

Whole genome sequencing of 223 individuals from 19 isolates (Fig. 1a; Supplementary Table 1) revealed 23,868,644 SNPs with a genome-wide distribution of 1 SNP per 9.94 bp on average.

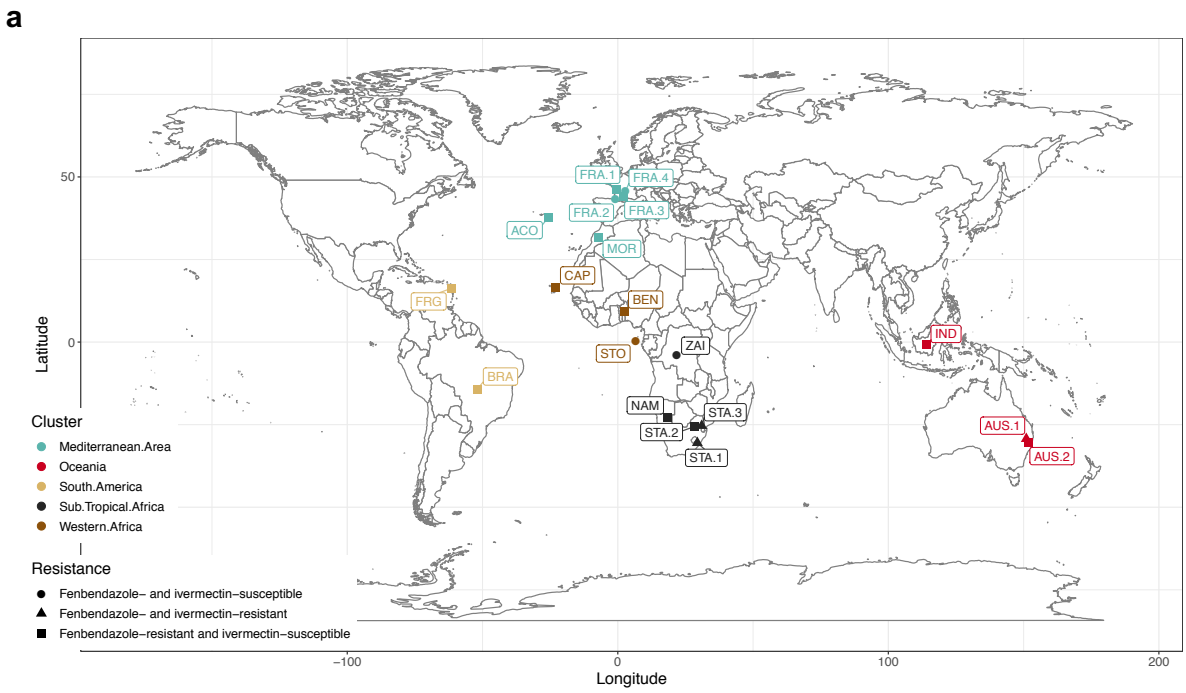


Figure 1a. Global distribution of *Haemonchus contortus* isolates. Isolates sampled are coloured by geographical region (sand: South-America, brown: Western-Africa, dark green: Mediterranean area, black: Subtropical Africa, red: Australia). Shape indicates the anthelmintic resistance status of each isolate to fenbendazole (resistant =squares; susceptible =circles) or resistant to both ivermectin and fenbendazole (triangles).

103

104 Only a proportion of the filtered SNPs were called in more than half the individuals (n
105 = 3,338,155 SNPs), with only 411,574 SNPs segregating with a MAF > 5% in those
106 individuals. Estimates of nucleotide diversity (π) in isolates with at least five
107 individuals ranged from 0.44% (STA.2) to 1.3% (NAM; Supplementary Table 2).
108 Variance in π across autosomes among isolates was partly explained by isolate
109 mean coverage ($F_{(1,84)} = 9.49$, $P = 0.003$; Supplementary note). To account for this
110 bias, estimates were obtained from three subsets of isolates with the highest
111 coverage (greater than 8× on average) from France (FRA.1 and FRA.2),
112 Guadeloupe (GUA) and Namibia (NAM) that yielded slightly higher values that
113 ranged between 0.65 and 1.14%. Overall, these data show equivalent diversity
114 levels to *Drosophila melanogaster* (ranging between 0.53% and 1.71%)[16] but
115 represented approximately 1.6 to 45-fold greater diversity than two filarial nematode
116 species for which similar statistics are available (0.02% for *Wuchereria bancrofti*
117 larvae[17]; 0.01% and 0.4% for π_S and π_N in *Onchocerca volvulus*[18] Supplementary
118 Fig. 1).

119

120 To begin to explore the global diversity of *H. contortus*, we performed a principal
121 component analysis (PCA) of genetic variation, which revealed three broad genetic
122 clusters of isolates (Fig. 1b) that largely coincided with the geographic region from
123 which they were sampled, including: (i) Sub-tropical African isolates (NAM, STA and
124 ZAI), (ii) Atlantic isolates including Morocco (MOR), São Tomé (STO), Benin (BEN),
125 Brazil (BRA) and Guadeloupe (GUA), and (iii) the remainder from the Mediterranean
126 area (FRA, ACO) and Oceania (AUS.1, AUS.2 and IND). The greatest diversity

among samples was identified in the African and South American isolates, with East African samples spread along PC1 and West African and South American samples along PC2.

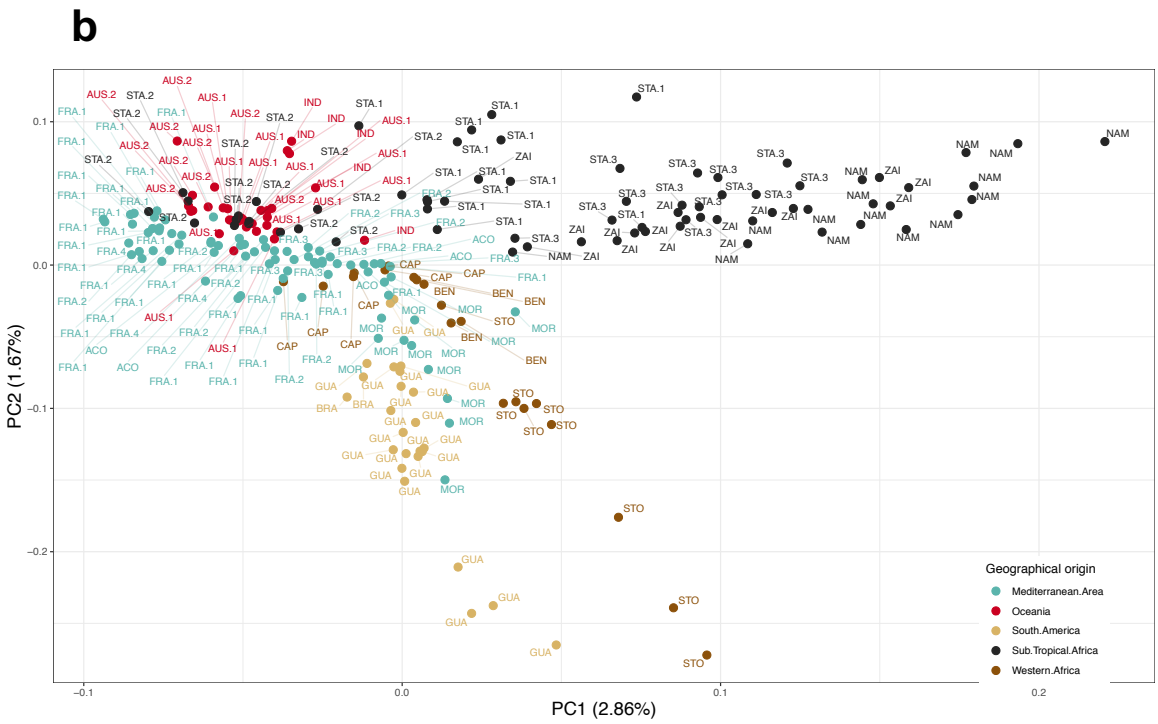


Figure 1b. Principal component analysis based on genotype likelihood inferred from whole genome sequences of 223 individual males (243,012 variants considered). Samples are coloured by geographic region described in panel 1a.

Using estimates of nucleotide diversity estimates, together with the the *C. elegans*[20] mutation rate and assuming a balanced sex ratio[21], we inferred the current effective population size (N_e) of *H. contortus* to be between 0.60 and 1.05 million . MSMC analysis, which models past recombination events based on heterozygosity patterns along the genome, revealed the historical N_e has remained within a slightly lower

range of values for most of the sampled time interval, i.e. from 2.5 to 500 thousand years ago (kya), with extreme estimates falling between 1.5 x 10^5 and 6.1 x 10^5 individuals for GUA (633 years ago) and NAM (415 years ago) isolates, respectively (Fig. 1c). *Ne* estimates remained relatively constant in most populations until 2.5 Kya; since this time, GUA and FRA.1 isolates suffered a more drastic reduction in N_e, which ranged for each isolate between 0.859 x 10^5 and 1.2 x 10^5 individuals approximately 633 and 452 years ago, respectively (Fig. 1c). (Fig. 1c).

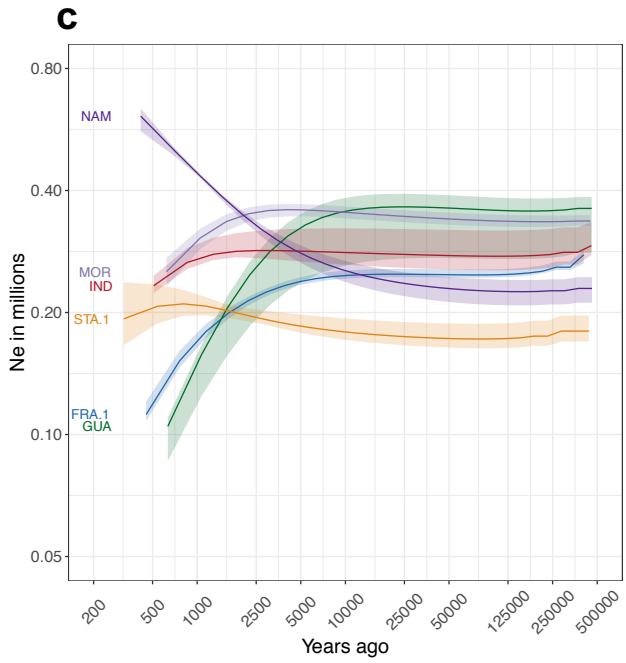


Figure 1c. MSMC effective population size across time for 6 isolates. Coloured shaded area represents range of values estimated from a cross-validation procedure with five replicates, computed by omitting one chromosome out at a time.

154 ***Global population connectivity of Haemonchus contortus is characterised by***

155 ***old and new migration***

156 To explore the global connectivity between isolates, we used a number of

157 complementary approaches. Phylogenetic relationships determined by nuclear

158 (Supplementary Fig. 2) and mitochondrial (Fig. 2a; PCA of mtDNA genetic diversity

159 is presented in Supplementary Fig. 3) diversity broadly supported the initial PCA

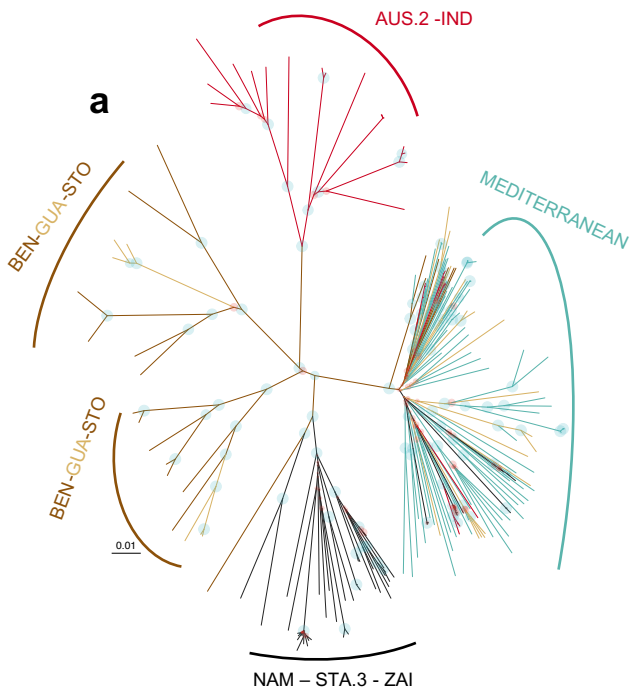
160 analysis (Fig. 1b), each revealing three main groups of samples partitioned by broad

161 geographic regions, *i.e.* Africa, Oceania and Mediterranean groups. However, the

162 mitochondrial data revealed further subdivision of the Oceanian and Mediterranean

163 clades than nuclear data alone.

164



165

166 **Figure 2a.** Unrooted maximum likelihood phylogenetic tree constructed constructed

167 using 3,052 SNPs from 223 individual mitochondrial genomes. Circles indicate

168 bootstrap support for each branch, blue if support was higher than 70%, red

elsewhere. Branches leading to a sample identifier are coloured by geographic region described in Figure 1a. Mitogroups are annotated with constitutive sample populations.

The presence of close genetic relationships between geographically distant isolates, 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 differentiation (measured by F_{ST}) between geographically distant French (FRA) and Oceanian (AUS.1, AUS.2, IND) isolates (Fig. 2b).

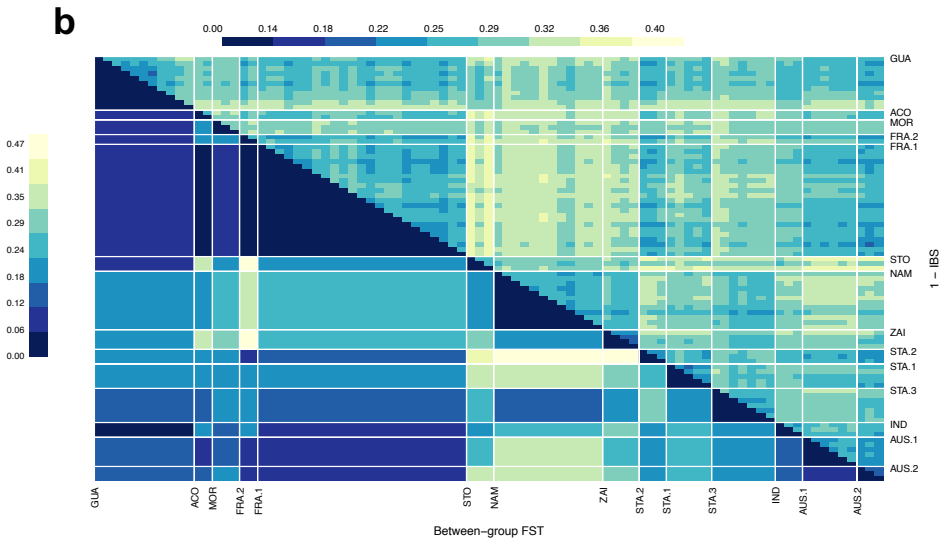
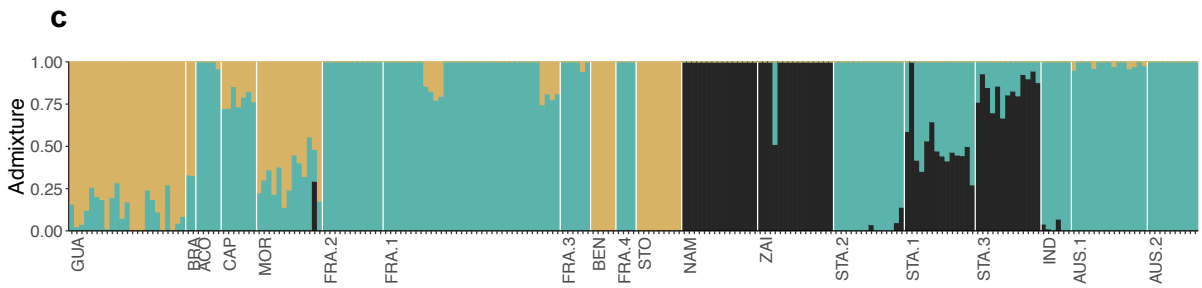


Figure 2b. Matrix showing pairwise dissimilarity between individuals (upper right) and pairwise F_{ST} between isolates (lower left).

The greatest genetic dissimilarity was found between subtropical African isolates and the French and Australian isolates, with mean F_{ST} across comparisons of 0.21 (F_{ST}

186 range according to isolate pairs: 0.06 - 0.42) and 0.24 (F_{ST} range: 0.17 - 0.33)
187 respectively (Fig. 2b, Supplementary Fig. 4). The divergence of African isolates was
188 likely due to the higher within-isolate diversity (+14% higher mitochondrial nucleotide
189 diversity) relative to others ($P = 0.004$; mitochondrial nucleotide diversity of 0.63% ±
190 0.37%, 0.58% ± 0.31%, 0.66 ± 0.32%, 0.6% ± 0.35%, for Mediterranean, Oceanian,
191 American and south-African isolates respectively).
192
193 The higher genetic differentiation (8.9% difference, $F_{(1,134)} = 11.36$, $P = 0.0009$)
194 between STO and other isolates relative to other pairwise comparisons, almost
195 certainly reflects the isolation of an island population relative to continental
196 populations (Fig. 2b). STO samples also displayed higher genetic dissimilarity (5.7%
197 difference, $F_{(1,7567)} = 581$, $P < 10^{-4}$) to other other samples. Inference of the joint
198 history of STO and other African populations supported this view. The best-fitting
199 model (supplementary Table 3) was consistent with either ancient symmetrical gene
200 flow followed by isolation or an early split followed by secondary contact before
201 isolation with MOR population (supplementary Table 3). This latter demography
202 would underpin the pattern of admixture detected between STO and MOR (Fig. 2c).
203



204

205 **Figure 2c**. Admixture analysis of 223 individuals. A cluster size of K = 3 is
206 presented, determined from sites with minor allele frequency above 5% and call rate
207 higher than 50%. Admixture pattern for other values of K is provided in
208 supplementary Figure 5. Isolates are presented sorted along their longitudinal range,
209 and samples sorted by assignment to the 3 clusters.
210
211 A closer inspection of GUA samples revealed a mixed ancestry, likely derived from
212 West African and Mediterraneanheritage. A subset of GUA samples showed limited
213 genetic differentiation to FRA isolates (F_{ST} range = 0.07 - 0.10; Fig. 2b), whereas the
214 remaining GUA samples were genetically similar to isolates from the West African
215 coast, *i.e.* ACO, BEN, MOR, STO (Fig. 2b & c), as indicated by lower F_{ST} estimates
216 with these isolates (F_{ST} range = 0.07 - 0.13; Wilcoxon's test, $P = 0.07$; Fig. 2b) and
217 evidence of shared ancestry from the admixture analysis (Fig. 2c, Supplementary
218 Fig. 5 and 6). A particularly close relationship was identified between STO and GUA
219 ($F_{ST} = 0.10$; Fig. 2b). This conflicting genetic origin among GUA sub-isolates is
220 responsible for the higher nucleotide diversity observed in GUA as a whole
221 (Supplementary Table 2).
222
223 The patterns of genetic connectivity support at least three distinct migration events in
224 time and space that we investigated using forward genetic simulations
225 (supplementary Table 3). First, we detect an out-of-Africa scenario, whereby the
226 greatest diversity was sampled within Africa, and that isolates outside of Africa
227 represent a subset of this diversity. Consistent with this hypothesis, nuclear genome
228 variation of non-African isolates experienced a genetic bottleneck that occurred

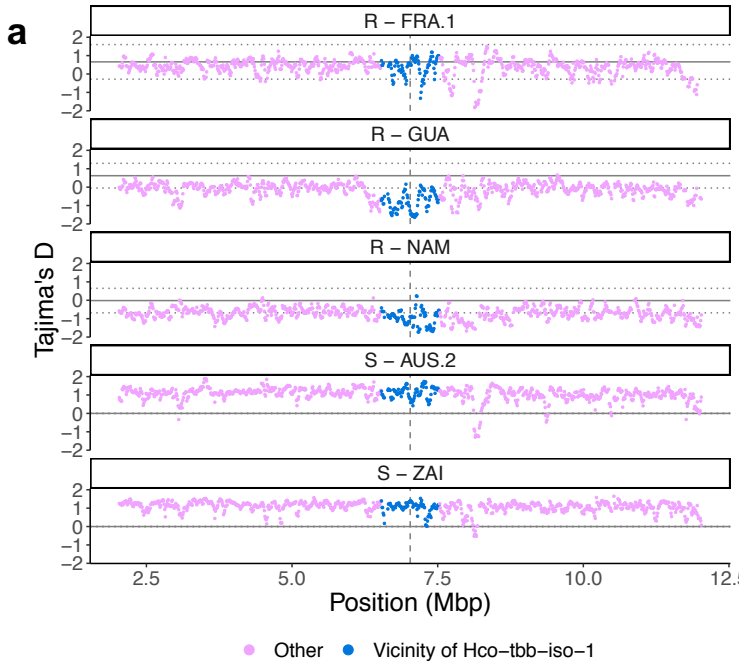
229 between 2.5 and 10 kya that is not present in the African isolates sampled (Fig. 1c).
230 Bayesian coalescent estimate of this initial divergence from mitochondrial genome
231 data yielded an overlapping time range of between 3.6 kya and 4.1 kya
232 (Supplementary Fig. 7). Genetic simulations of the joint demography between FRA.1
233 and African populations also favoured complex scenarios involving early split
234 (maximum likelihood estimates ranging between 13 and 25 Kya) followed by ongoing
235 gene flow with STA populations or more recent isolation with Namibia
236 (supplementary Table 3). Secondly, the genetic connectivity of West African and
237 American isolates is consistent with parasites spreading during the trans-Atlantic
238 slave trade movement. The scenario linking GUA and STO was compatible with
239 initial population division occurring around 1640 Common Era (CE) ± 167 years
240 (supplementary table 3), consistent with migration associated with colonization and
241 slave trade that occurred in the West Indies under French influence during that
242 time[22]. The third pattern of connectivity likely reflects British colonisation of Australia
243 in late 1700's: the interweaving of Australian and South-African worms into the
244 Mediterranean phylogenetic haplogroup (Fig. 2b) may mirror the foundation of
245 Australian Merino sheep, which were first introduced into Australia from South Africa,
246 before additional contributions from Europe and America were made[23,24]. The
247 admixture pattern observed for worms from the two countries matched their shared
248 ancestry (Fig. 2c), as well as a genetic connection between America and Australia
249 (seen for the AUS.1 isolate; Fig. 2c). Although maximum likelihood estimates
250 supported these scenarios with the isolation of European and South-African isolates
251 occurring between 1794 and 1871 CE (supplementary Table 3), broad uncertainty
252 limited our ability to specifically define the timings of these events (Supplementary

253 Table 3). However, the split between Australian isolates occurred in 1895 CE ± 132
254 years, consistent with the initial foundation of the sheep industry in this
255 country.These complex patterns reiterate that human movement has played an
256 important role in shaping the diversity of this livestock parasite throughout the world.
257

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

260 The extensive use of anthelmintics has been and remains the primary means of *H.*
261 *contortus* and other gastrointestinal control worldwide. This strategy has resulted in
262 the independent emergence of drug-resistant isolates throughout the world, which
263 now limits farming in some areas. Strong selection should impact the distribution of
264 genetic variation within isolates; signatures of selection may reveal genes associated
265 with drug resistance, knowledge of which may contribute to monitoring the
266 emergence and spread of drug resistant isolates, and the design of new control
267 strategies.
268 The genetic determinants of benzimidazole resistance is perhaps the best
269 characterised of all anthelmintics, with either one of three amino acid residue
270 changes - F167Y, E198A, and F200Y - in the beta tubulin isotype 1 (*Hco-tbb-iso-1)*
271 protein capable of mediating phenotypic resistance. We identified indications of a
272 selective sweep in the region surrounding the *Hco-tbb-iso-1* locus in the resistant
273 isolates analysed. Focusing on three isolates with the highest coverage, we found an
274 average 2.31-fold reduction of Tajima's *D* coefficient within 1 Mbp (Fig. 3a,
275 Supplementary Fig. 8) and an average 33% reduction in nucleotide diversity
276 (Supplementary Fig. 9) within this region relative to the rest of chromosome I.

277



278

279 **Figure 3a**. Analysis of Tajima's *D* surrounding the *Hco-tbb-iso-1* locus on

280 chromosome I. A total of 10-Mbp surrounding was considered (pink), of which 1 Mbp

281 nearest to *Hco-tbb-iso-1* is highlighted (blue). Isolates compared included

282 benzimidazole-resistant French (FRA.1), Guadeloupian (GUA), Namibian (NAM) or

283 benzimidazole susceptible Australian (AUS.2) and Zaire (ZAI) isolates. Mean

284 expected Tajima's *D* (solid grey line) and 99% confidence interval (dotted line) were

285 estimated from a 1,000 simulated 10-Kbp wide sequences following MSMC-inferred

286 demography.

287

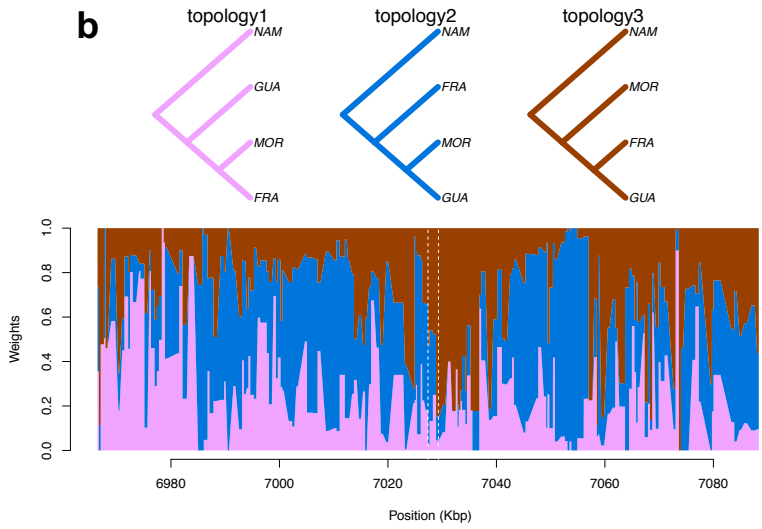
288 This signature was most evident in the GUA and NAM isolates, but was weaker in

289 the French isolate (FRA.1) due to both phenotypically susceptible and resistant

290 individuals being present (Supplementary Table 4). Phased genotypic information

291 over the whole *Hco-tbb-iso-1* (Supplementary Fig. 10) revealed that French and

292 Guadeloupian individuals had little divergence in their haplotypes, suggesting that
293 gene flow of resistant haplotypes between mainland France and the West Indies had
294 occurred (Supplementary Fig. 10). A topology analysis of a 100 Kbp region spanning
295 the *Hco-tbb-iso-1* locus supported the shared phylogenetic origin between FRA.1
296 and FRG (topology 3), whereas the surrounding region was in favour of topologies
297 congruent with overall population structure (Fig. 3b). This finding is consistent when
298 either Moroccan (Fig. 3b) or São Tomé populations are used as a population with
299 close ancestry with FRG (Supplementary Fig. 11).
300



301

302 **Figure 3b**. Topology weighing analysis of a 100-Kbp window centred on *Hco-tbb-*
303 *iso-1*. At each position, the weight of each of the three possible topologies inferred
304 from 50 Kbp-windows are overlaid. Topology 2 (blue) corresponds to an isolation-by-
305 distance history, while topology 3 (brown) would agree with shared introgressed
306 material between worm populations from French mainland into Guadeloupe. The
307 *Hco-tbb-iso-1* locus is indicated by vertical dashed lines.
308

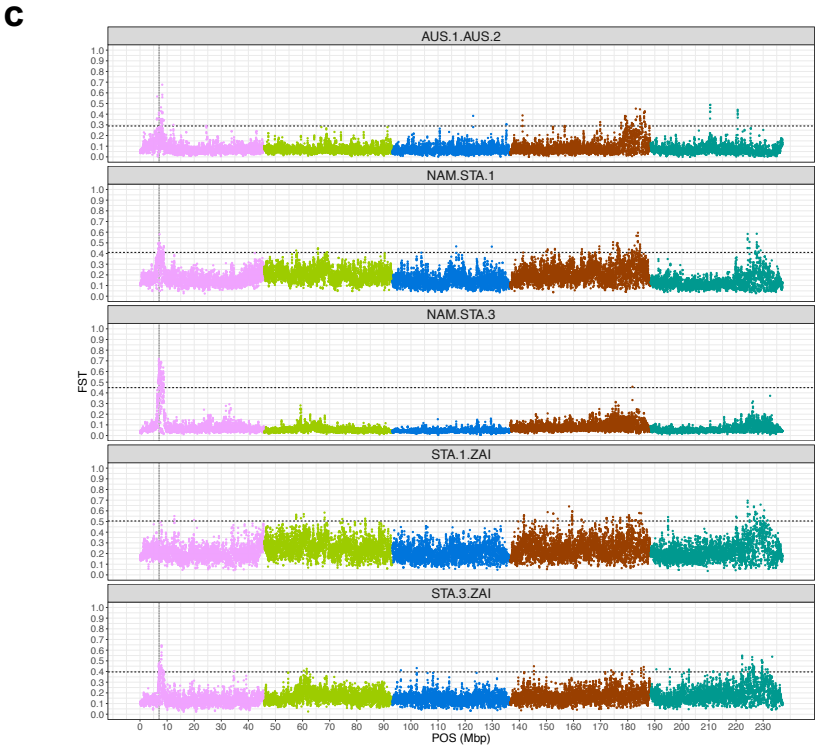
309 We analysed the frequency of the three well-characterised resistance-associated
310 mutations affecting codon positions 167 (T to A), 198 (A to T) and 200 (T to A)
311 (Supplementary Table 4; Supplementary Table 5; Supplementary Fig. 12 & 13).
312 The F200Y homozygous genotype was the most common and widespread resistant
313 genotype (n = 39), and accounted for all samples from the Guadeloupe (n = 14) and
314 the White-River South-African (STA.3; n = 6) isolates. Variants at codons 167 and
315 198 were much less common, i.e. 13 mutant allelic carriers were observed at
316 position 198 and only one F167Y mutant was identified in a French isolate. No
317 double homozygous mutants were found, but three individuals from France, Australia
318 and South-Africa were heterozygous at both positions 198 and 200. In each case,
319 inspection of the sequencing reads revealed that the two mutations never appeared
320 together in the same sequencing read, suggesting they cannot co-occur in *cis* (on
321 the same chromosome copy). Genotype frequencies were consistent with isolate-
322 level benzimidazole efficacies as measured by the percentage of egg excretion after
323 treatment (Supplementary Table 5). Samples analysed from suspected susceptible
324 isolates always presented with the susceptible genotype for each of the three
325 positions considered (n = 20).
326
327 The genetic determinants of ivermectin resistance are largely unknown, but many
328 genes have been proposed to be associated with resistance; although one
329 interpretation of this observation is that ivermectin resistance is a multigenic trait, a
330 major locus associated with resistance in Australian and South African *H. contortus*
331 has recently been mapped to a region approximately 37-40 Mbp along
332 chromosome V [12]. To examine for the presence of ivermectin-mediated selection in

333 our data, a pairwise differentiation scan was performed between known ivermectin-

334 resistant isolates and other isolates sharing same genetic ancestry, *i.e.* STA.1 and

335 STA.3 versus NAM and ZAI in Africa, AUS.1 against AUS.2 in Australia (Fig. 3c).

336



337

338 **Figure 3c**. Genome-wide differentiation scan between pairs of ivermectin-resistant

339 (AUS.1, STA.1, STA.3) and populations with no evidence of ivermectin resistance.

340 Chromosomes are coloured and ordered by name, *i.e.* from I (pink) to V (forest

341 green). Horizontal line represents the 0.5% F_{ST} quantile cut-off. Vertical line on

342 chromosome I points at *Hco-tbb-iso-1* locus.

343

344 The previously described region on chromosome V[12] appeared differentiated,

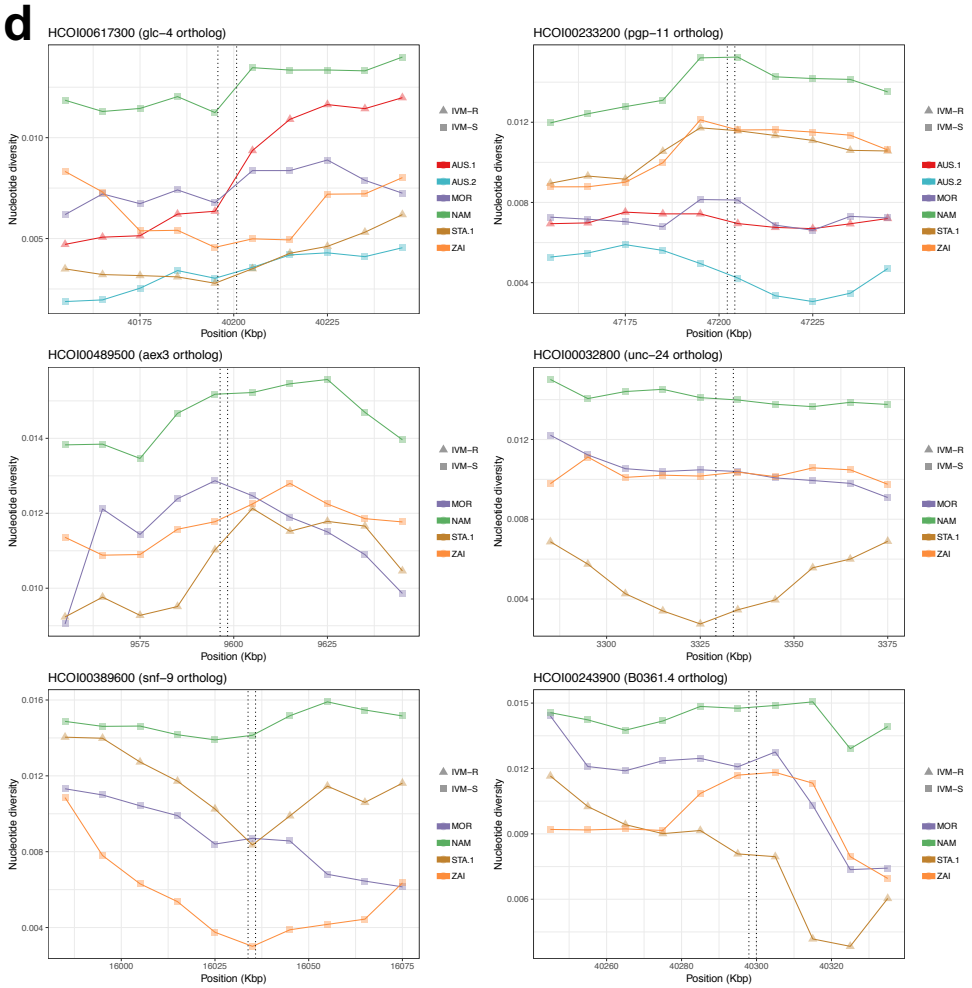
345 particularly between pairwise comparisons involving the South-African resistant

346 isolate (Fig. 3c). A second major differentiation hotspot spanned a 3 Mbp region of

347 chromosome I and encompassed the *Hco-tbb-iso-1* locus (Fig. 3c). Although non-

348 synonymous mutations at codon positions 167, 198 and 200 of the β-tubulin isotype
349 1 are associated with resistance to benzimidazoles, it has been proposed that there
350 may also be an association between this gene and ivermectin resistance[25,26].
351 Although our data seem to support this association, an attempt to narrow-down the
352 differentiation signal in this region found significant differentiation in only two 10-kbp
353 windows in two comparisons (NAM vs STA.1 and NAM vs. STA.3, and these two
354 windows did not overlap the *Hco-tbb-iso-1* locus. Furthermore, the fact that all of our
355 ivermectin-resistant isolates were also benzimidazole-resistant suggests that this
356 signal is confounded; pairwise F_{ST} analyses between ivermectin-resistant and -
357 susceptible isolates from the field will always be biased toward strong differentiation
358 around the *Hco-tbb-iso-1* locus due to loss of diversity in benzimidazole-resistant
359 isolates. As such, it was not possible to confirm the putative association between the
360 *Hco-tbb-iso-1 locus* and ivermectin resistance. We note that the lack of QTL
361 evidence in this region from controlled genetic crosses using ivermectin selection[12]
362 supports the conclusion that standing genetic variation at the *Hco-tbb-iso-1*locus is
363 unlikely to be directly influenced by ivermectin.
364
365 To further investigate more subtle signatures of selection, we computed the XP-CLR
366 coefficient, which simultaneously exploits within-isolate departure from neutrality and
367 between-isolate allele-frequency differences [27]. This analysis relies on called
368 genotypes rather than genotype likelihoods, but is robust to SNP uncertainty[27].
369 Within continent pairwise comparisons of African (NAM, MOR, STA.1, STA.3, ZAI)
370 and Australian isolates (AUS.1, AUS.2) yielded 1,740 hotspots of diversification,
371 48% of which were contained within a gene locus (Supplementary Fig. 14,

372 Supplementary Table 6). Among these hotspots, two known candidate genes
373 associated with ivermectin resistance were identified, namely an ivermectin sensitive
374 glutamate-gated chloride channel[28] (*HCOI00617300*, *glc-4* ortholog) on chromosome
375 I, and a P-glycoprotein coding gene already involved in ivermectin susceptibility in
376 the equine ascarid *Parascaris sp* [29], (*HCOI00233200, pgp-11* ortholog) on
377 chromosome V. While the former showed indication of reduced genetic diversity in
378 its vicinity in resistant isolates relative to others (0.48% ± 0.09% difference in
379 nucleotide diversity between the two groups, $t = 5.25$, $P < 10^{-4}$; Fig. 3d), the genetic
380 diversity pattern in the 100 Kbp window surrounding the latter was similar across
381 isolates ($t = 0.004$, $P = 0.99$; Fig. 3d), suggesting its role may not be specifically
382 related to ivermectin resistance.
383



384

385 **Figure 3d**. Nucleotide diversity over 100 Kbp windows surrounding positional

386 candidate genes for ivermectin resistance. Candidate gene position is highlighted by

387 vertical dashed lines. Ivermectin-resistant isolates (IVM-R) appear as triangles and

388 squares indicate susceptible isolates (IVM-S).

389

390 GO term enrichment analysis of all XP-CLR significant regions identified 26

391 significant terms (Supplementary Table 7). The top ten most significant GO terms

392 encompassed enzymatic-related activity, neurotransmitter transporter activity

393 (GO:0005326, $P = 5.8 \times 10^{-4}$) and response to drug (GO:0042493, $P = 4.3 \times 10^{-3}$).

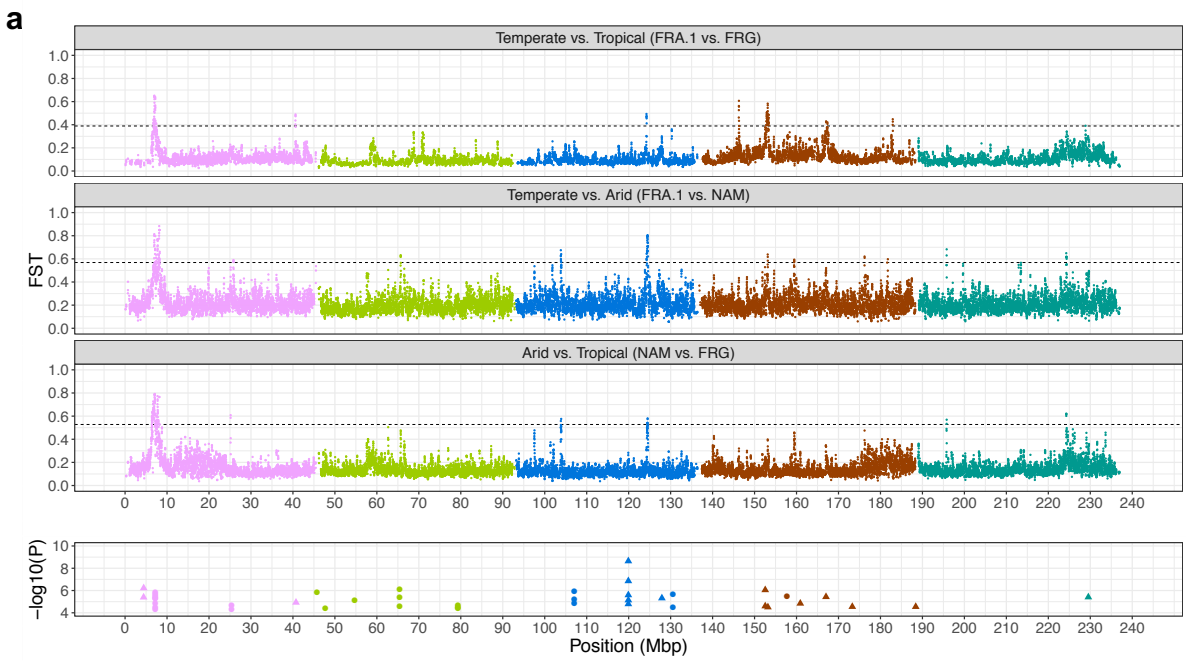
394 Additional significant biological process associated terms were related to phenotypes

395 tightly linked to anthelmintic effects. For example, pharyngeal pumping[30]
396 (GO:0043051, $P = 5.8 \times 10^{-3}$) and oviposition (GO:0046662, $P = 8.6 \times 10^{-3}$) genes
397 were enriched, both of which are phenotypes linked to ivermectin and its effect on
398 parasite fecundity[31]. Neurotransmission is the primary target of anthelmintics such as
399 macrocyclic lactones[32] and levamisole[33]; genes with GO terms related to
400 neurotransmission (GO:0001505, $P = 9.1 \times 10^{-3}$) were significantly enriched across
401 comparisons. Anthelmintic-associated GO terms including "response to drug",
402 "regulation of neurotransmitter" and "neurotransmitter transporter activity" were
403 significantly enriched in every comparison involving a South-African multi-resistant
404 field isolates (STA.1, Supplementary Table 8). Four candidate genes
405 (*HCOI00389600*, *HCOI00032800*, *HCOI00243900*, *HCOI00489500*) were found
406 overlapping XP-CLR signals of selection and thus seem candidates for drug
407 resistance loci in *H. contortus*, as supported by the functions of their *C. elegans*
408 orthologs (*snf-9*, *unc-24*, *B0361.4*, *aex-3* respectively). The *snf-9* gene encodes a
409 neurotransmitter:sodium symporter belonging to a family of proteins involved in
410 neurotransmitter reuptake, and the latter three genes are expressed in neurons.
411 *Unc-24* and *B0361.4* are involved in response to lipophilic compounds and *aex-3* is
412 critical in synaptic vesicle release[34]. A reduction in nucleotide diversity surrounding
413 *unc-24*, *B0361.4*, *aex-3* in the STA.1 isolate in comparison to others (Fig. 3d) may
414 reflect evidence of drug selection.
415
416

417 ***Climatic adaptation has shaped genomic variation between isolates***

418 In addition to putative anthelmintic-related GO terms, response to stress
419 (GO:0006950, $P = 5.9 \times 10^{-3}$) was among the top ten significant biological process
420 ontologies associated with genes under diversifying selection. Although anthelmintic
421 exposure would be associated with significant stress on susceptible (and perhaps
422 resistant or tolerant) parasites, all free-living stages of *H. contortus* will be exposed
423 to and must tolerate abiotic factors such as temperature or humidity prior to infection
424 of a new host. To evaluate the impact of such climatic stressors, a genome scan for
425 genetic differentiation between isolates categorised by by the climactic conditions
426 prevailing at their sampling locations, i.e. arid (Namibia), temperate (France
427 mainland) and tropical (Guadeloupe), was performed (Fig. 4a).

428



429

430 **Figure 4a. Genomic signal of climate adaptation**

431 Top three panels represent F_{ST} values plotted against genomic position (Mbp),
432 colored and ordered by chromosome name (from I to V). Horizontal dashed line

433 represents the 0.5% quantile. Bottom panel shows genomic positions of significant
434 associations between annual precipitation (circle) and temperature annual range
435 (triangle).
436
437 A major signal of differentiation formed of two windows (6.925 – 6.995 Mbp and
438 7.165 - 7.205 Mbp on chromosome I) was shared by all pairwise comparisons. Six
439 genes were found within these two regions (Supplementary Table 9), among which
440 orthologs of *C. elegans* genes *hprt-1*, *cpb-1, B0205.4* were identified. A highly
441 differentiated 260 Kbp region of chromosome III also repeatedly occurred across
442 comparisons. A window between 31.155 Mbp and 31.185 Mbp was common to
443 comparisons involving isolates from arid areas (NAM vs. FRA.1 and NAM vs. GUA),
444 and a second window (31.345 Mbp to 31.415 Mbp) was common to comparisons
445 between tropical isolates and others (GUA *vs*. FRA.1 and GUA *vs*. NAM). While the
446 two annotated genes in the latter region were not associated with any biological
447 description, the first window overlapped a chromo-domain containing protein
448 (*HCOI01540500*; Supplementary Table 89). This gene is an ortholog of *Pc*
449 (*FBgn0003042*) in *D. melanogaster*, a chromo-domain subunit of the Polycomb
450 PRC1-complex that specifically recognizes trimethylated lysine7 of histone 3
451 (H3K27me3)[35]. Nucleotide diversity over these genomic windows revealed reduced
452 genetic diversity in FRA.1 relative to isolates from hotter climates (Fig. 4b). The
453 contrasted diversity downstream from *HCOI01540500* (31.225 Kbp to 31.350 Kbp,
454 Fig. 4b) found between GUA and NAM certainly contributed to higher F_{ST} in these
455 windows. Additional differentiation analyses performed at the gene level with base-
456 pair resolution highlighted a few discrete locations of elevated F_{ST} common to every

457 comparison and overlapping intron 4 and exon 5 of this gene (Fig. 4b). Translated

458 consensus exon 5 sequences revealed the highest divergence in the GUA isolate

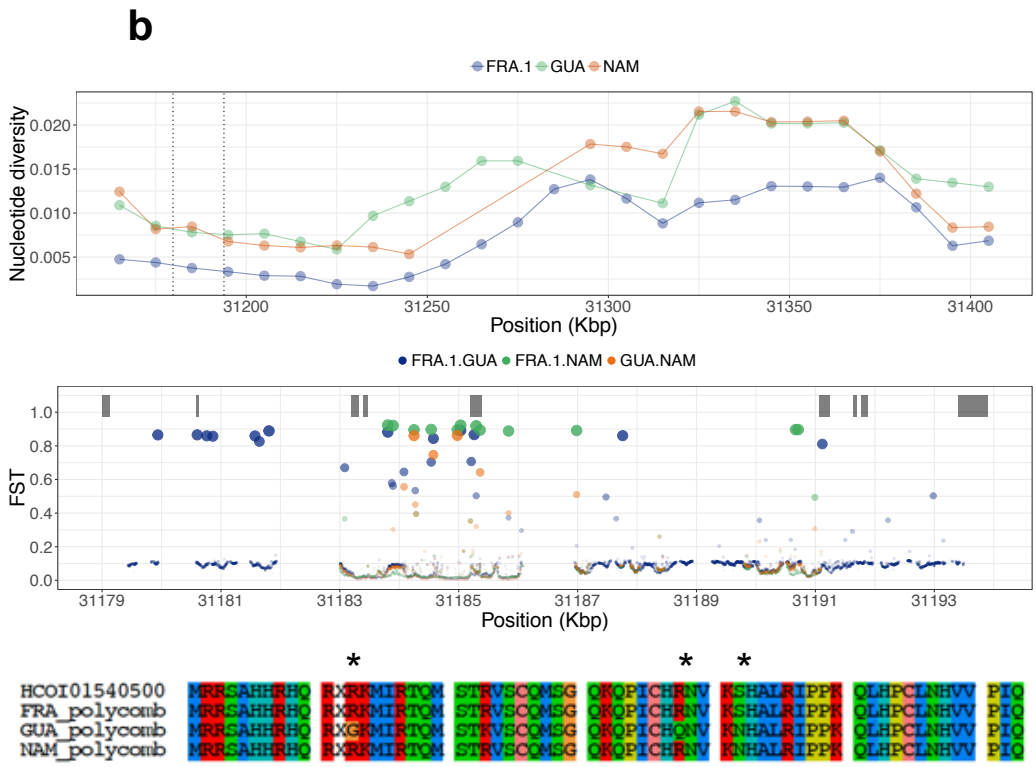
459 (92.1% identity with reference sequence; supplementary Fig. 15), characterized by

460 multiple amino acid changes whose putative functional consequences remain

461 unknown.

462

463



464

465 **Figure 4b. A *Polycomb group protein coding* gene underpins major**

466 **differentiation signal between populations from arid, temperate and tropical**

467 **areas.**

468 Top panel shows nucleotide diversity estimates for 10 Kbp genomic windows

469 spanning windows of chromosome III with high genetic differentiation for climate.

470 Dashed lines indicate the *Pc* ortholog boundaries. Within these boundaries, F_{ST}

471 estimates along every base-pair of the gene sequence are represented below (circle
472 size proportional to F_{ST} coefficient), with predicted exon model shown as grey
473 rectangles. The bottom panel shows translated consensus sequences of exon 5 for
474 populations of interest with asterisks marking mutations.
475
476
477 To further explore the impact of climatic conditions of genetic diversity, we used a
478 random forest based statistical approach to quantify the relationship between genetic
479 information along environmental gradients, and relative impact of specific bioclimatic
480 variables on genetic diversity. Bioclimatic variables were derived from monthly
481 temperature and precipitation records, with the aim to represent annual trends or
482 seasonality over the 1970 to 2000 time period (Supplementary Fig. 16;
483 Supplementary Table 10). This analysis randomly samples subsets of sites (encoded
484 as SNP frequencies) that can be partitioned into groups based on differences in
485 climactic variables to estimate the predictive ability of climatic variables. Annual
486 precipitation (BIO12), and temperature annual range (BIO7) were revealed to be the
487 most important bioclimatic variables impacting genetic variation (Fig. 4c).
488

b

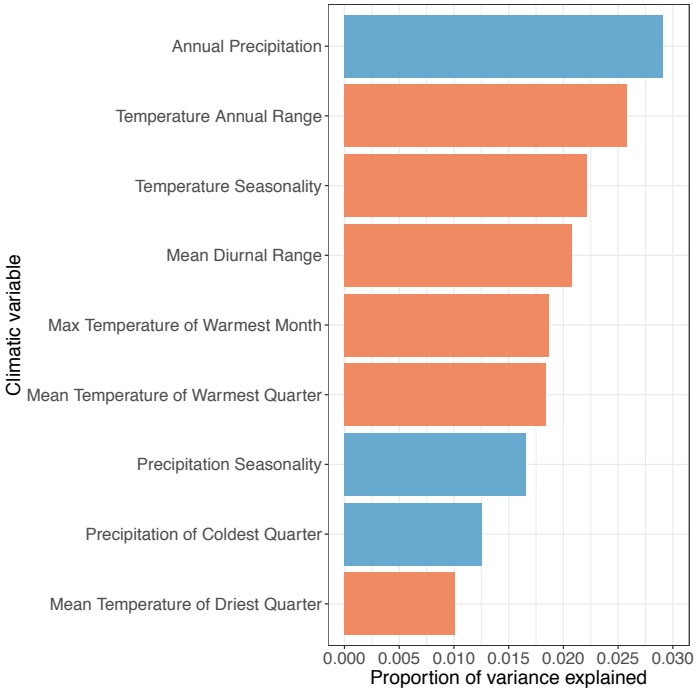


Figure 4c. Proportion of genetic variance explained by precipitation- (blue) and temperature-associated (orange) climatic variables. Variants from eight isolates with no record of ivermectin resistance (AUS.2, FRA.1, GUA, IND, MOR, NAM, STO, ZAI) were analysed against a set of nine variables (from a total of 19), selected to minimize correlation between variables. Annual precipitation (BIO12) and temperature annual range (BIO7) are main contributors of genetic variance.

To identify genes that might be impacted by these variables, we performed a genome-wide test for association between SNP variants, and each of these two environmental variables, accounting for genetic structure between isolates. 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 11). 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

504 13) associations with BIO12 fell within a 6 Kbp window (7,177,538 bp and 7,183,617
505 bp), overlapping the region universally differentiated between climatic areas.
506 Chromosome III contained 11 significant associations for both BIO7 (n = 6) and
507 BIO12 (n = 5). The two most significant associations were also found on this
508 chromosome for BIO7 (positions 26,824,240 and 26,824,197 bp; $P = 6.34 \times 10^{-9}$ and
509 $P = 1.39 \times 10^{-7}$, respectively). These two SNPs fell within the *HCOI00198200* locus,
510 which codes for a metallopeptidase M1, an ortholog of the *C. elegans* gene *anp-1*.
511 Additional BIO7-associated SNPs were found within (16,544,889 bp) and in the
512 vicinity (15,905,141 & 15,905,191 bp) of a solute carrier coding gene
513 (*HCOI00312500*) on chromosome IV, whose orthologs in *C. elegans* are under the
514 control of *daf-12*, a key player in dauer formation. Of note, two *H. contortus* orthologs
515 of components of the dauer pathway in *C. elegans*, namely *tax-4* (*HCOI00661100*)
516 and *daf-36* (*HCOI00015800*), were among genes under diversifying selection
517 (Supplementary Fig. 14), but neither of these genes contained SNPs identified as
518 being associated with the two most significant bioclimatic variables.

519 # Discussion

520 The ecology and epidemiology of gastrointestinal nematodes have been well
521 characterized and exploited to build mathematical models to guide treatment
522 decision in the field[36]. On the contrary, knowledge about their genetic diversity
523 remains limited ; a better understanding of population structure and selective
524 pressure applied by environmental factors would yield better predictive power of their
525 range dispersion[37]. By exploiting a broad collection of samples from globally-
526 distributed isolates, together with chromosome-scale assembly and extensive
527 individual resequencing, we have performed an in-depth characterization of *H.*
528 *contortus* genetic diversity, explored historical contributions to its current population
529 structure and identified important drivers shaping the genome of this parasite.
530
531 *H. contortus* populations displayed high levels of nucleotide diversity, consistent with
532 early estimates based on mitochondrial data[39,40], and recent re-sequencing
533 experiments of inbred isolates[12]. These extreme levels of genetic diversity are
534 thought to arise from both a large census population[42] and the high fecundity of *H.*
535 *contortus* females[21].
536
537 An early attempt based on a set of genome-wide AFLP markers obtained for 150
538 individual worms from 14 countries supported the first exploration of *H. contortus*
539 population structure at the continent level [40]. Analyses of these data identified three
540 to four (Africa, South-East Asia, America and Europe) main phylogenetic clusters as
541 well as evidence for the strong genetic connectivity between Australian, South-
542 African and European isolates[40]. Our genome wide data corroborated these early

543 results. However, the use of a chromosome-scale assembly and individual
544 resequencing contributed to identify genome-wide patterns of genetic diversity in its
545 chromosomal context, and in turn, provided sufficient resolution to identify genes
546 likely associated with selective advantage against drug or climate selection
547 pressures. This had not been possible in previous attempts with AFLP markers [40].
548
549 Major past human migrations and associated sheep movements have contributed to
550 the mixing of parasite populations, which partly accounts for the limited genetic
551 structure and extensive admixture between some of our globally distributed isolates.
552 Our data supported an out of Africa expansion derived from ancestral populations
553 from Western Africa, with a bottleneck dated back between 2.5 to 10 kya. Sheep
554 domestication originally took place in the Middle East around 10 kya[43], before
555 introduction into Eastern Africa and subsequent spread, likely through cultural
556 diffusion, toward southern Africa approximately 2 to 2.5 kya [44]. The timing of the
557 bottleneck identified in our data and occurring between 2.5 and 10 kya is compatible
558 with major migrations of pastoralist populations that ultimately resulted in the import
559 of small ruminants into Southern Africa [44]. The simultaneous increase in rainfall in
560 central Africa around 10.5 kya[45] would have supported the population expansion and
561 dispersal of *H. contortus*. In addition, early radiation towards Asia observed in our
562 data was congruent with evidence obtained from the study of retroviral insertions
563 within the sheep genome that suggested direct migration between Africa and
564 southwest Asia[46], and with the timing of Asian sheep expansion between 1.8 and 14
565 kya[47].
566

567 The genetic congruence of parasites in Guadeloupe with those from both West
568 Africa and the Mediterranean region is consistent with co-transportation of
569 livestock, including sheep, during the discovery and subsequent colonization of
570 America.. Woolly Churra sheep were originally brought to the Caribbean by Spanish
571 conquistadors [49,50], before West African breeds more suited to tropical conditions
572 were transported , resulting in an admixed Carribean sheep population [49,50]. The
573 timing of genetic admixture here overlaps with centuries of human and presumably
574 livestock movement during the transport of slaves, most of whom originated in
575 West Africa and were transported to colonies that included the French West
576 Indies[22]. The Mediterranean ancestry of *H. contortus* isolates in Guadeloupe and the
577 close relationship with French worm populations suggest additional sheep transport
578 occurred between French mainland and its Guadeloupian overseas territory.
579 Although these movements are difficult to track precisely, live sheep were shipped
580 aboard slave trade vessels departing from French harbours[51], and may have
581 introduced European *H. contortus* to the island. Based on historical data[49,50], a
582 Spanish lineage of *H. contortus* on Guadeloupe would also be expected to exist,
583 though data are missing to confirm this. It can be speculated that the shared
584 ancestry between Guadeloupian and Moroccan worm isolates might result from the
585 introduction of Spanish Churra sheep, as this breed emerged in Spain while the
586 region was under Arab influence between the 8th and 13th century[49,52]. Introduction of
587 *H. contortus*-infected sheep from the Maghreb may have been associated with the
588 spreading of African worms in Spain, and ultimately, Guadeloupe.
589

590 The widespread use of anthelmintic drugs has been a major selective force shaping
591 standing genetic variation in the isolates analysed. Adaptation to drug exposure
592 was clearly illustrated in the genetic diversity surrounding the β-tubulin coding gene,
593 a target of benzimidazole drugs; strong loss of diversity was observed at this locus
594 in many distinct isolates phenotypically characterised to be resistant to this class of
595 drug. The co-occurrence of the same resistant haplotypes in geographically
596 disconnected isolates is almost certainly due to the independent evolution of
597 benzimidazole resistance as has been described previously[53]; however, some
598 evidence of shared ancestry between French and Guadeloupe resistant individuals
599 emphasises the risk of spreading resistance without careful monitoring of parasite
600 populations during livestock trade. Furthermore, this highlights the risk of evolution
601 of resistance in other parasitic nematodes, for example, parasites of medical
602 interest that are treated with benzimidazoles in mass-drug administration
603 programs[54].
604
605 Ivermectin is an important anthelmintic for parasite control in both veterinary and
606 medical settings[55]. Worldwide emergence of ivermectin-resistant veterinary parasitic
607 nematodes[5] and evidence of reduced efficacy in human filarial nematodes[56]
608 underline the urgency and importance of a better understanding of the mechanisms
609 involved[55]. Our data identified genes previously associated with ivermectin
610 resistance in parasitic nematodes of either veterinary (*pgp-11 in Parascaris sp.*[29]) or
611 medical (*aex-3* in *Onchocerca volvulus*[57]) interest and uncovered new candidates.
612 Additional validation study will determine whether these could serve as that may be
613 useful markers in the field to monitor drug efficacy.. Of note, support for a major QTL

614 for ivermectin resistance on chromosome V recently identified in an introgression
615 experiment in two resistant isolates[38] was also present in our genome-wide data and
616 warrants further investigation.
617
618 Our broad sampling throughout the global range of *H. contortus* has enabled the first
619 analyses into climate-driven adaptation in a parasitic nematode. Understanding how
620 climate shapes parasite genetic variation is of primary importance to foresee
621 consequences of climate change on parasite phenology and range dispersion.
622 Parasite dispersal is largely driven by their hosts but *H. contortus* free-living stages
623 experience climatic conditions that affect their development[2] and constrain their
624 spatio-temporal dispersal[58]. Observations in Northern Europe suggest that climate
625 change has already altered *H. contortus* winter phenology [59]. Our results suggest
626 that adaptation toward annual precipitation was mostly under the control of variation
627 on chromosome I, however, no obvious candidate genes could be identified. In
628 addition, the *Hco-tbb-iso-1* locus was close to the region of interest, and linked
629 variation in allelic frequency at this locus as a result of benzimidazole selection
630 cannot be ruled out. A second region of chromosome III was associated with both
631 temperature- and precipitation-related variables, within which biologically relevant
632 genes could be identified; first, the strongest genetic associations with annual
633 temperature range were found in a metallopeptidase with zinc ion binding function
634 (*anp-1* ortholog), an enzyme linked to drought stress tolerance in *Drosophila* [60], and
635 second, an ortholog of *Pc* displayed strong genetic differentiation between arid- and
636 wetter temperate or tropical environmental conditions. This gene has been linked to
637 putative epigenetic regulation of xeric adaptation in *Drosophila melanogaster*, where

638 *Pc* mutants display lower resistance to desiccation stress[61]. This finding is also
639 corroborated by observations in plants showing Polycomb-mediated regulation of
640 climate-induced phenotypes[62]. Further links to climatic adaptation include
641 identification members of the dauer pathway[63]; *tax-4* and *daf-36* orthologs were
642 under diversifying selection and a *daf-12*-respondent solute carrier was associated
643 with annual temperature range. Dauer is a developmentally arrested stage in *C.*
644 *elegans* that is triggered by environmental stress and mediates tolerance to
645 unfavourable conditions until better conditions are met[63], and can occur in parasitic
646 nematodes like *H. contortus* under semi-arid conditions[64]. Evidence for climatic
647 adaptation suggests that adaptation in the face of climate change will be
648 constrained by available genetic variability at temperature-selected loci, that may
649 both limit or enable range expansion depending on the region. By a better
650 understanding of the interaction between climatic conditions and phenotypes such
651 as hypobiosis (a temporary developmental arrest during unfavorable conditions),
652 optimisation of treatment timing may be possible to maximise control efficacy.
653
654 In summary, our data describes the extensive global and genome-wide diversity of
655 the blood-feeding parasitic nematode *H. contortus*, and how this diversity has been
656 shaped by adaptation to its environment and to drug exposure. Understanding the
657 mechanism(s) by which parasites adapt to fluctuating environmental conditions both
658 within and outside their hosts will have important implications for field management
659 of parasitic nematodes in both veterinary and medical settings. Further
660 characterisation of these putative strategies, together with genetic covariation of drug
661 resistance genes, should contribute to refining epidemiological models and guide

662 treatment decision-trees for a more sustainable management of worm populations in

663 the face of a changing climate.

664 # Legends to Figures

665 ***Figure 1. Global diversity of Haemonchus contortus.***

666 (a) Global distribution of *Haemonchus contortus* isolates. Isolates sampled are
667 coloured by geographical region (sand: South-America, brown: Western-Africa, dark
668 green: Mediterranean area, black: Subtropical Africa, red: Australia). Shape indicates
669 the anthelmintic resistance status of each isolate to fenbendazole (resistant
670 =squares; susceptible =circles) or resistant to both ivermectin and fenbendazole
671 (triangles).
672 (b) Principal component analysis based on genotype likelihood inferred from whole
673 genome sequences of 223 individual males (243,012 variants considered). Samples
674 are coloured by geographic region described in panel 1a.
675 (c) MSMC effective population size across time for 6 isolates. Coloured shaded area
676 represents range of values estimated from a cross-validation procedure with five
677 replicates, computed by omitting one chromosome at a time.
678

679 ***Figure 2. Global connectivity of Haemonchus contortus isolates.***

680 (a) Unrooted maximum likelihood phylogenetic tree constructed using 3,052 SNPs
681 from 223 individual mitochondrial genomes. Circles indicate bootstrap support for
682 each branch, blue if support was higher than 70%, red elsewhere. Branches leading
683 to a sample identifier are coloured by geographic region described in Figure 1a.
684 Mitogroups are annotated with constitutive sample populations.

685 (b) Matrix showing pairwise dissimilarity between individuals (upper right) and
686 pairwise F_{ST} between isolates (lower left).
687 (c) Admixture analysis of 223 individuals. A cluster size of K = 3 is presented,
688 determined from sites with minor allele frequency above 5% and call rate higher than
689 50%. Admixture pattern for other values of K is provided in supplementary Figure 5.
690 Isolates are presented sorted along their longitudinal range, and samples sorted by
691 assignment to the 3 clusters.
692

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

694 (**a**) Analysis of Tajima's *D* surrounding the *Hco-tbb-iso-1* locus on chromosome I. A
695 total of 10-Mbp surrounding was considered (pink), of which 1 Mbp nearest to *Hco-*
696 *tbb-iso-1* is highlighted (blue). Isolates compared included benzimidazole-resistant
697 French (FRA.1), Guadeloupian (GUA), Namibian (NAM) or benzimidazole
698 susceptible Australian (AUS.2) and Zaire (ZAI) isolates. Mean expected Tajima's D
699 (solid grey line) and 99% confidence interval (dotted line) were estimated from a
700 1,000 simulated 10-Kbp wide sequences following MSMC-inferred demography.
701 (**b**) Topology weighing analysis of a 100-Kbp window centred on *Hco-tbb-iso-1*. At
702 each position, the weight of each of the three possible topologies inferred from 50
703 Kbp-windows are overlaid. Topology 2 (blue) corresponds to an isolation-by-distance
704 history, while topology 3 (brown) would agree with shared introgressed material
705 between worm isolates from French mainland into Guadeloupe. The *Hco-tbb-iso-1*
706 locus is indicated by vertical dashed lines.
707 (**c**) Genome-wide differentiation scan between pairs of ivermectin-resistant (AUS.1,
708 STA.1, STA.3) and isolates with no evidence for ivermectin resistance.

709 Chromosomes are coloured and ordered by name, i.e. from I (pink) to V (forest
710 green). Horizontal line represents the 0.5% F_{ST} quantile cut-off. Vertical line on
711 chromosome I points at *Hco-tbb-iso-1* locus.
712 (**d**) Nucleotide diversity over 100 Kbp windows surrounding positional candidate
713 genes for ivermectin resistance. Candidate gene position is highlighted by vertical
714 dashed lines. Ivermectin-resistant isolates (IVM-R) appear as triangles and squares
715 indicate susceptible isolates (IVM-S).

716 ***Figure 4. Genomic signal of climate adaptation***
717 (**a**) Top three panels represent F_{ST} values plotted against genomic position (Mbp),
718 colored and ordered by chromosome name (from I to V). Horizontal dashed line
719 represents the 0.5% quantile. Bottom panel shows genomic positions of significant
720 associations between annual precipitation (circle) and temperature annual range
721 (triangle).
722 (**b**) A *Polycomb group protein coding* gene underpins major differentiation signal
723 between populations from arid, temperate and tropical areas. Top panel shows
724 nucleotide diversity estimates for 10 Kbp genomic windows spanning windows of
725 chromosome III with high genetic differentiation for climate. Dashed lines indicate the
726 *Pc* ortholog boundaries. Within these boundaries, F_{ST} estimates along every base-
727 pair of the gene sequence are represented below (circle size proportional to F_{ST}
728 coefficient), with predicted exon model shown as grey rectangles. The bottom panel
729 shows translated consensus sequences of exon 5 for populations of interest with
730 asterisks marking mutations.
731 (**c**) Proportion of genetic variance explained by precipitation- (blue) and
732 temperature-associated (orange) climatic variables. Variants from eight isolates with

733 no record of ivermectin resistance (AUS.2, FRA.1, GUA, IND, MOR, NAM, STO,
734 ZAI) were analysed against a set of nine variables (from a total of 19), selected to
735 minimize correlation between variables. Annual precipitation (BIO12) and
736 temperature annual range (BIO7) are main contributors of genetic variance.
737
738

739 # Materials and methods

740 ***Sample DNA extraction and sequencing***

741 A total of 267 individual male *H. contortus* were obtained from a collection held at
742 INRA[16] (metadata for all samples is presented in detail in Supplementary Table 1).
743 The sampling regime was motivated to delineate the contribution of major
744 evolutionary forces, *i.e.* migration and selection (drug and climate) but also
745 constrained by the material available in the collection. Because migration was likely
746 to match human history, isolates from western African countries and southern
747 America were selected to address the contribution of slave trade history to the
748 structuring of *H. contortus* populations; isolates from former colonies of the British
749 Empire (South-Africa, Australia) were sampled to establish the connectivity between
750 worm populations from Europe and these countries. Ivermectin-resistant isolates
751 (AUS.1, STA.1, STA.3) were also retained to evaluate how anthelmintics had shaped
752 *H. contortus* genomic variability (drug efficacy data have been provided in
753 supplementary Table 4). Isolates were selected based on available material to
754 ensure minimal sample size (n = 9) per isolate and proper allele frequency
755 estimation. Following these criteria, 19 isolates from 12 countries were available
756 (Supplementary Table 1). Note that the second Australian population (AUS.2) was
757 obtained from an Italian laboratory (labelled ITA_NAP, supplementary Table 1).
758 Samples were gathered between 1995 and 2011 (Supplementary Table 1), and
759 stored in liquid nitrogen upon collection. Four isolates were fenbendazole susceptible
760 (Fig. 1 and supplementary table 4; triangles; FRA.2, FRA.4, STO, ZAI).
761

762 DNA was extracted with the NucleoSpin Tissue XS kit (Macherey-Nagel GmbH&Co,
763 France) following the manufacturer's instruction. Sequencing libraries were prepared
764 as previously described[12]. DNA libraries were sequenced with 125 bp paired-end
765 reads on an Illumina Hiseq2500 platform using V4 chemistry (Supplementary Table
766 1). A second round of sequencing was performed with 75 bp paired-end reads to
767 increase the coverage of 43 samples (Supporting Fig. 15). After sequencing, two
768 samples were identified to be heavily contaminated by kraken-0.10.6-a2d113dc8f[65]
769 and were discarded. In total, 18 sequencing lanes consisting of 4,152,170,256 reads
770 were sequenced, the raw data of which are archived under the ENA study accession
771 PRJEB9837.

772 ***Sequencing data processing***

773 Read mapping to both the mitochondrial and nuclear genomes (v3.0, available at
774 ftp://ngs.sanger.ac.uk/production/pathogens/Haemonchus_contortus) was performed
775 using SMALT (http://www.sanger.ac.uk/science/tools/smalt-0) with a median insert
776 size of 500 bp, k-mer length of 13 bp, and a stringency of 90%. For samples that had
777 two or more BAM files (when split across multiple sequencing lanes), the BAM files
778 were merged using samtools v.0.1.19-44428[66], and duplicated reads removed using
779 Picard v.2_14_0 (https://github.com/broadinstitute/picard) before performing
780 realignment around indels using Genome Analysis Toolkit (GATK v3.6)[67]
781 RealignerTargetCreator. Mean coverage of the mitochondrial and genomic genomes
782 were estimated using GATK DepthOfCoverage, revealing coverage lower than the
783 estimated target coverage (original 8x) for most samples. Individuals with more than
784 80% of their mitochondrial sequence with at least 15 reads and a mean

785 mitochondrial genome coverage of at least 20× were retained for population genetic
786 inferences (n = 223 individuals).

787 ***Nuclear genome SNP calling***

788 To call SNPs, we used GATK HaplotypeCaller in GVCF mode, followed by joint
789 genotyping across samples (GenotypeGVCFs) and extraction of variants
790 (SelectVariants), resulting in a total of 30,040,159 unfiltered SNPs across the five
791 autosomes. Sex determination in *H. contortus* is based on an XX/XO system; as only
792 male worms were sequenced, their hemizygous X chromosome would have revealed
793 limited phylogenetic information relative to the autosomes, and was henceforth
794 excluded from further analysis.
795
796 Low coverage sequencing will inadvertently bias allele sampling at heterozygous
797 SNPs, resulting in excess homozygous genotypes particularly if stringent filtering is
798 applied during SNP calling. To circumvent this issue, we applied the GATK Variant
799 Quality Score Recalibration
800 (https://gatkforums.broadinstitute.org/gatk/discussion/39/variant-quality-score-
801 recalibration-vqsr), which first uses a reference (truth) SNP set to estimate the
802 covariance between called SNP quality score annotations and SNP probabilities,
803 followed by application of these probabilities to the raw SNPs of interest.
804
805 The reference "truth" SNP database was generated from the intersection of variants
806 called from samples with at least a mean of 10× coverage (n = 13) using three
807 independent SNP callers: (i) samtools mpileup (-q20 –Q20 –C50 -uD), (ii)
808 Freebayes[68] v.9.9.2-13-gad718f9-dirty (--min-mapping-quality 20 --min-alternate-

809 count 5 --no-indels --min-alternate-qsum 40 --pvar 0.0001 --use-mapping-quality --
810 posterior-integration-limits 1,3 --genotype-variant-threshold 4 --use-mapping-quality -
811 -site-selection-max-iterations 3 --genotyping-max-iterations 25 --max-complex-gap
812 3), and (iii) GATK HaplotypeCaller followed by hard filtering (--QD<2, --DP>10000, --
813 FS>60, --MQ<40, --MQRS <-12.5, --RPRS<-8). The three variant call sets were
814 merged (GATK CombineVariants with --genotypeMergeOptions UNIQUIFY),
815 resulting in an intersecting set of 794,606 SNPs (extracted with GATK
816 SelectVariants). The GATK VariantRecalibrator model was trained with this
817 reference SNP database with a 90% prior likelihood, before being subsequently
818 applied to the raw set of SNPs (n = 30,040,159). The estimation was run for several
819 truth sensitivity threshold values ranging from 90 to 99.9%. After visual inspection of
820 the additional number of SNPs by using sensitivity tranche curves (Supplementary
821 Fig. 16a), a 97% sensitivity threshold was applied to the raw SNP set with GATK
822 ApplyRecalibration, resulting in a total set of 23,868,644 SNPs spanning the five
823 autosomes. Variant depth of coverage (DP) and strand bias (FS) were the main
824 drivers of SNP removal (Supplementary Fig. 16b).
825
826 Called SNPs were used for particular analyses (*Ne* trajectory through time, cross-
827 population composite likelihood-ratio) that could not be performed under the
828 probabilistic framework that relied on genotype likelihoods as implemented in
829 ANGSD[69] v. 0.919-20-gb988fab (Supplementary note). These data were also used
830 to estimate average differentiation between isolates. However, within isolate diversity
831 and admixture were analysed using ANGSD[69] (Supplementary note).

832 ***Mitochondrial DNA data processing***

833 The mitochondrial genome exhibited an average coverage depth of 322× (ranging
834 from 24x to 5,868x) per sample (Supplementary Table 1). Mitochondrial reads were
835 extracted and filtered from poorly mapped reads using samtools view (-q 20 -f
836 0x0002 -F 0x0004 -F 0x0008) and from duplicated reads using Picard v.2_14_0
837 MarkDuplicates (https://github.com/broadinstitute/picard). Realignment around indels
838 was applied with GATK[70] and SNP were subsequently called using samtools[71]
839 mpileup using only reads that achieved a mapping quality of 30 and base quality of
840 30. The occurrence of heterozygous sites in mtDNA, known as heteroplasmy, has
841 been described across vertebrate species[72] and in other nematode species,
842 including *C. briggsae*[73] and *C. elegans*[74]. However, heterozygous sites may also
843 occur as technical artefacts as a result of genetically similar sequences shared
844 between the nuclear and mitochondrial genomes, *i.e.*, numts[75]. To exclude sites
845 prone to heterozygous signals from further phylogenetic inference analysis, a SNP
846 calling procedure was implemented with the HaplotypeCaller tool to apply hard
847 filtering parameters on the raw SNP sets (QD>=10, FS<=35 MQ>=30
848 MQRankSum>=-12.5 and ReadPosRankSum >=-8) with a minimum depth of 20
849 reads. This procedure excluded 1,354 putative heterozygous sites, and retained 72%
850 of the putative SNP sites (3,052 out of 4,234 SNPs). Nucleotide diversity and
851 Tajima's *D* were computed by sliding-windows of 100 bp using vcftools v.0.1.15[76]. A
852 principal component analysis (PCA) was performed on genotypes using the
853 SNPrelate package[77] in R version 3.5[78]. A consensus fasta sequence was
854 subsequently generated with GATK FastaAlternateReferenceMaker for each sample
855 using the filtered variant set, which was used for the phylogenetic analyses.

856 ***Diversity and divergence analysis***

857 Genome-wide nucleotide diversity (π) was computed for each isolate with at least
858 five individuals using ANGSD[69]. Using genotype likelihoods (GLs) from samtools[71]
859 (option GL=1) as an input, variants were included that had a minimal supporting
860 evidence of 5 reads, and base and mapping quality phred scores of at least 20. As π
861 values were biased by population mean coverage (Supplementary Fig. 17), π was
862 also calculated from a subset of isolates containing individuals with a minimum mean
863 coverage of 5×: this was limited to France (FRA.1, n = 5, mean coverage of 7.66×),
864 Guadeloupe (n = 5, mean coverage of 12.75×) and Namibia (n = 6, mean coverage
865 of 9.85×).

866

867 F_{ST} was estimated from the VQSR-called genotypes between isolates with at least
868 five individuals using the Weir-Cockerham estimator[79] in vcftools v0.1.15[76]. To
869 prevent artifactual signal linked to variation in coverage between isolates, F_{ST} was
870 calculated on subsets of SNPs, binned based on their minor allele frequency (MAF)
871 in 10% increments. The maximum F_{ST} value calculated was retained for comparison
872 (Supplementary Fig. 4) and the maximal value was considered as reported
873 elsewhere[80]. The resulting F_{ST} estimates were not biased by coverage, as measured
874 by negligible correlation between pairwise F_{ST} coefficients and associated population
875 cross-coverage (Pearson's $r_{(136)} = -0.05$, $P = 0.55$; Supplementary Fig. 18).

876

877 The pairwise sequence divergence between individual samples was calculated using
878 the Hamming distance, i.e. 1-IBS, using PLINK[81], considering the only individuals
879 with a mean coverage above 2.5× as failure to do so yielded biased estimates

880 (Supplementary note, Supplementary Fig. 19). A neighbour-joining tree of Hamming
881 distances calculated from the nuclear DNA genotypes was built using the R package
882 *ape*[82].
883
884 PCA on genotypes inferred from genotype likelihoods of the 223 samples was
885 generated using ANGSD ngsCovar, filtering for sites with base and mapping quality
886 phred scores less than 30, minimum depth of 5 reads, and a SNP *p*-value (as
887 computed by ANGSD) below 10^{-3}. Clustering was robust to coverage variation and
888 closely matched the PCA from the VQSR-called SNP genotypes (Supplementary
889 Fig. 20).

890 ***Phylogenetic inference***

891 To determine the phylogenetic structure of the cohort, the 223 consensus
892 mitochondrial fasta sequences were first aligned using Muscle v3.8.31[83], followed by
893 stringent trimming of sequence alignments using Gblocks[84]. The most likely
894 evolutionary model, GTR substitution model with rate heterogeneity modelled by a
895 gamma distribution with invariable sites, was determined using modelgenerator
896 v.0.851[85]. A maximum-likelihood tree was subsequently generated using PhyML[86]
897 v.20120412, with branch supports computed using 100 bootstraps.

898 ***Admixture analysis***

899 Admixture was determined using NGSAdmix[87]. This tool relies on genotype
900 likelihoods to account for data uncertainty, and has been shown to produce robust
901 inferences about population ancestry from low coverage samples alone, or a mixture
902 of low and higher coverage samples[87]. This analysis was performed on 223

903 samples, for *K* ranging from 2 to 10 clusters (Supplementary Fig. 5 and 6), retaining
904 sites with less than 50% missing data across individuals and minor allele frequency
905 (MAF) above 5%. Five iterations were run omitting one autosome out at a time, and
906 the best K was chosen as the first value that would minimize the median absolute
907 deviation across runs (Supplementary Fig. 6). Sample coverage did not affect the
908 results (Supplementary Fig. 21).

909 ***Effective population size and population divergence dating***

910 The effective population size (*Ne*) trajectory through time, and the cross-divergence
911 time between populations, were estimated using MSMC2[88,89]. This approach uses
912 patterns of heterozygosity along the genome to identify past recombination events
913 modelled as Markov processes[90]. Mutation density along the sequence mirrors either
914 recent (long tract of limited diversity) or older (enrichment in heterozygosity over
915 short distances) events. According to coalescent theory[91], at any given time, the
916 amount of recombination is proportional to *Ne*.
917
918 MSMC2 was applied to individuals with a mean coverage above 10×, limiting the
919 analysis to six isolates (FRA.1, GUA, IND, MOR, NAM, STA.1), and considering four
920 haplotypes per isolate. Beagle v4.1 was used to impute missing genotypes and to
921 establish phase in VQSR SNP calls. Imputation accuracy analysis revealed a 7.2%
922 and 9.0% discordance rates at the individual and site levels respectively
923 (Supplementary Fig. 22). Input files were created following MSMC recommendations
924 and available msmc-tools (https://github.com/stschiff/msmc-tools). Briefly, the
925 reference fasta sequence was masked with SNPable
926 (http://lh3lh3.users.sourceforge.net/snpable.shtml) to extract regions of unambiguous

927 read mapping in chromosome-specific bed files (using the available
928 msmc_create_map_mask.py python script). Negative bed files indicating regions
929 with sufficient coverage at the individual level were created from samples bam files
930 with the bamCaller.py script and filter out sites with coverage below genome-wide
931 average depth. Finally, MSMC2 input files were created for each chromosome with
932 the generate_multihetsep.py script and concatenated into a single input file. For
933 each isolate, estimates were averaged across five runs, leaving one chromosome
934 out at a time for cross-validation, using rho/mu parameter value of 6.22 (average
935 recombination rate of 1.68 cM/Mbp[12] and considering a mutation rate similar to that
936 of *C. elegans* mutation rate, 2.7 x 10^{-9} per site per generation[20]). MSMC2 times and
937 coalescent rates were scaled to real time and population sizes by assuming the
938 same mutation rate[20], a balanced sex-ratio[21] and an inferred generation interval of
939 40 days (the sum of 10 days to reach mature free-living infective larvae from the egg
940 stage, and a 30-day prepatent period for fully mature egg-laying females)[92].
941
942 Migratory scenarios between populations were determined using the forward
943 simulation framework implemented in δaδi[93]. Under a given evolutionary scenario,
944 this software models the expected joint site frequency spectrum between multiple
945 populations using a diffusion equation. These expected values are then used to
946 compute the most likely demographic parameters knowing the observed site
947 frequency spectrum. For each model, four rounds of forward simulations were run
948 with 10, 20, 30 and 40 replicates respectively using previously published python
949 scripts[94] (https://github.com/dportik/dadi_pipeline). Model Akaike Information
950 Criterion (AIC) were compared for ranking scenarios, from which the lower the

951 score, the more likely the outcome.

952 We first compared a divergence scenario without migration against models including
953 symmetrical and asymmetrical gene flow before isolation. In case migration was the
954 most likely, more complex models (involving split with ancestral (a)symmetrical gene
955 flow, with or without population size change, or models involving secondary contact
956 with/without gene flow and population size change) were tested. However, initial
957 exploration indicated a likely lack of power in our design to accurately estimate
958 parameters of more complex demographic models than the split and isolation model.
959 Nevertheless, these models still provide the most likely scenario and their output
960 have been listed in supplementary Table 3.

961 Parameters were scaled to real time using same parameters as for MSMC2
962 inference. Standard deviations of timing estimates for the simple split and isolation
963 models were obtained using the Godambe Information Matrix[95] applied to 100
964 simulated site frequency spectra produced with the *ms* software[96] under the most
965 likely demographic model.

966

967 Additional support to the estimates from the nuclear genome were obtained from a
968 phylogenetic analysis of coding sequences using BEASTv1.10[97]. Mitochondrial
969 coding sequences were extracted from the consensus sequence of every individual,
970 concatenated per individual, and aligned using Musclev3.8.31[83]. A Bayesian skyline
971 model[98] was used, with a HKY substitution model and a strict clock model, as other
972 modalities yielded weak effective sample size (ESS) and unstable parameter values.
973 Clock rate was set to *C. elegans* mitochondrial mutation rate, i.e. 1×10^{-7} per site per
974 generation[74], as variation in sampling date was not sufficient to estimate molecular

975 rate. Parameters showed sufficient sampling (effective size above 200) after
976 50,000,000 iterations and a burn-in of the first 20 million steps. Node ages were
977 scaled to years assuming a generation interval of 40 days. A maximum clade
978 credibility tree was generated with TreeAnnotator v1.10.1
979 (http://beast.community/treeannotator).

980 ***Diversifying selection scan***

981 To further characterize the genetic diversity in *H. contortus* populations, we identified
982 genomic regions under diversifying selection using XP-CLR[27]. This approach takes
983 advantage of both within-population distortion of the allele frequency spectrum, and
984 between-population differences in allele frequencies in the vicinity of selective
985 sweeps[27]. Although XP-CLR is robust to SNP ascertainment bias[27], the analysis was
986 restricted to isolates with at least five individuals with a minimum mean coverage of
987 2× per individual. Unphased VQSR-derived genotypes were filtered to retain SNPs
988 with a within-isolate call rate >80% and MAF >5%. The analysis was run on every
989 one of the 22 possible within-continent pairwise comparisons of the retained isolates
990 (Australia and Africa) with the following options: -w1 0.0001 500 2000 –p0 0. This fit
991 a grid of putatively selected points every 2 Kbp along the genome, with a sliding
992 window size around grid points of 0.01 cM, interpolating SNP position in Morgans
993 from the average recombination rate for each chromosome[12]. Down-sampling was
994 applied to windows where more than 500 SNPs were found to keep SNP numbers
995 comparable between regions, and no LD-based down-weighing of the CLR scores
996 was applied. A selection score was subsequently computed at every position as the
997 root mean square of XP-CLR coefficients, and the highest 0.1% of selection scores
998 were deemed significant, as reported elsewhere[80].

999 ***Analysis of β-tubulin isotype 1 (Hco-btub-iso-1) and the genetic architecture of***
1000 ***benzimidazole- and ivermectin-resistance***

1001 The genomic coordinates of *Hco-btub-iso-1* were determined by blasting the gene
1002 coding sequence from WormBase Parasite[99] against current genome assembly
1003 (BLASTN, e-value < 10^{-50}). The SNP positions associated with codons 167, 198 and
1004 200 was determined to be at 7027535, 7027755, and 7027758 bp, respectively,
1005 along chromosome I after aligning the *Hco-btub-iso-1* consensus sequence and
1006 published sequence[100] (GenBank accession FJ981629.1) using muscle v3.8.31[83].
1007 Genotype and GLs at these positions were determined with ANGSD[69]. Genotypes
1008 were only considered for GL >60% (n = 74) as coverage bias occurred otherwise
1009 (Supplementary Table 6 and Supplementary Fig. 12). For these samples, phased
1010 and imputed genotypes from VQSR SNPs spanning the *Hco-btub-iso-1* locus were
1011 used to compute pairwise number of allele differences. Selection in the vicinity of the
1012 *Hco-btub-iso-1* locus was assessed by computing Tajima's *D* with ANGSD[101]. For
1013 benzimidazole-resistant isolates, neutral state was built with *ms*[102] by simulating 10-
1014 Kbp wide isolate-specific sequences (n=1,000) following the same coalescent
1015 scenario as predicted by MSMC2 for chromosome I (considering a recombination
1016 rate of 1.83 cM/Mbp[12]). Suitable *ms* input parameters were derived from MSMC2
1017 output files using the msmc2ms.py script from msmc-tools. This approach was
1018 implemented for isolates with sufficient mean depth of coverage, i.e. three
1019 benzimidazole-resistant isolates. In case of susceptible isolates, the lack of
1020 significant departure in the *Hco-btub-iso-1* vicinity relative to the rest of chromosome
1021 I was tested.
1022

1023 The introgression of resistant haplotypes from France mainland into Guadeloupian
1024 worm isolates was tested by a topology weighting analysis implemented with
1025 TWISST[103]. This method computes phylogenetic trees between a set of isolates,
1026 using genetic information from short sliding windows (50 Kbp) and by sampling with
1027 replacement individuals from each isolate. At each window, a weight is subsequently
1028 computed for each of the possible tree topology, ultimately providing inference of
1029 introgression events in discrete locations where tree topology connects
1030 phylogenetically distant isolates. Analyses were run using individuals with sufficient
1031 mean depth of coverage (5× and more) from French (FRA.1; n = 5; mean depth =
1032 7.62×) and Guadeloupian (GUA; n = 5; mean depth = 12.75×) isolates, adding
1033 Namibian isolate (n = 2; mean depth = 10.9×) as an outgroup. A fourth isolate was
1034 chosen for its common ancestry with Guadeloupian isolates: first, an analysis was
1035 run with worms from São Tomé (n = 2; mean depth = 7.01×), followed by a second
1036 analysis using Moroccan samples (n = 2; mean depth = 10.6×), to ensure that
1037 evidence of introgression was consistent in both cases.
1038
1039 To investigate the genetic architecture of ivermectin resistance, a differentiation scan
1040 was run between ivermectin-resistant isolates and isolates of unknown status
1041 sharing same ancestry, i.e. STA.1 and STA.3 versus NAM and ZAI in Africa, AUS.1
1042 against AUS.2 in Australia. Windowed F_{ST} estimates[79] were calculated every 10 Kbp
1043 with a 1 Kbp overlap along the genome with ANGSD[69], retaining sites with minimal
1044 depth of 5 reads, mapping and base quality phred scores of at least 30, missing rate
1045 below 50%, and windows with at least 1000 sites. Genomic coordinates of the top
1046 0.5% most differentiated windows were extracted and analysed by BLASTN

1047 (minimum e-value = 10^{-50}) against the published V1 *H. contortus* assembly[13], and
1048 annotated gene identifiers were inferred from the corresponding GFF file from
1049 WormBase Parasite[99].

1050 ***Identification of SNPs under environmental selection***

1051 To identify SNPs putatively influenced by environmental selection, we first defined
1052 the climatic conditions of each isolate using the Köppen-Geiger classification[104], and
1053 inferred from isolate geographical coordinates[105]. Three isolates with the best
1054 coverage, i.e. Namibia, France, and Guadeloupe, were used to contrast dry,
1055 temperate and tropical conditions, respectively. We performed a genome-wide scan
1056 using pairwise comparisons of F_{ST} using ANGSD[69], of which test values greater than
1057 the 0.5% quantile in at least two comparisons were analysed further.
1058
1059 To further explore the contribution of environmental climatic variables on standing
1060 genetic variation, we applied a machine-learning gradient forest algorithm[106] to
1061 quantify changes in genetic variation (fit as population SNP MAF) along
1062 environmental gradients. The gradients consisted of 19 bioclimatic variables[107]
1063 summarizing rainfall and temperature information recorded between 1970 and
1064 2000(supplementary table 9). For each SNP, a random forest of 500 trees was
1065 grown. For each tree, bootstrapped SNP MAFs were regressed against a random
1066 subset of bioclimatic variables to determine the variable that best partitioned the
1067 data, thereby building the first node that partitions the data into two sets of
1068 homogeneous observations. Iterations follow to determine subsequent nodes by
1069 resampling a random subset of bioclimatic variables until no observations are left.
1070 The proportion of variance explained by each bioclimatic variable is then averaged

1071 across SNPs, and the function of SNP frequency modification along bioclimatic
1072 variable is built.
1073 The analysis was performed on isolates with at least 5 individuals, a mean coverage
1074 of 2×, and no record of ivermectin resistance (AUS.2, FRA.1, GUA, IND, MOR,
1075 NAM, STO, ZAI). SNP MAFs were estimated from genotype likelihoods with
1076 ANGSD, and subsequently filtered to ensure a within-isolate MAF >10%, at least
1077 90% within-isolate call-rate, and shared across the eight considered isolates,
1078 resulting in 3,758 SNPs retained for further analysis.
1079
1080 Environmental variables were highly correlated (Supplementary Fig. 15a), resulting
1081 in instability in the predictor's importance. To minimise this effect, the gradient forest
1082 analysis was restricted to the 11 environmental variables showing least redundancy
1083 as assessed by a PCA (Supplementary Fig. 15b). Pair-wise Euclidean distances
1084 between variables was computed from their respective coordinates on the first two
1085 PCA axes (supplementary Fig. 15a). We selected variables with higher distances
1086 (mean distance > 0.85) with any others (BIO4, 7, 2, 15, 5, 10). Other variables
1087 defined three clusters (Supplementary Fig. 15b). Within each cluster, we picked the
1088 variable with closest distance from every other in the cluster, i.e. the variable
1089 summarizing others' contributions (BIO9, BIO13, BIO19). For the BIO12, 13, 16, 18
1090 cluster, we chose to pick BIO12 (annual precipitation) which is more relevant toward
1091 parasite life-cycle across climatic areas. Under temperate areas, quarter-based
1092 (BIO16, BIO18) or wettest month (BIO13) statistics would match seasons where
1093 hosts are housed. SNP-environment associations were further investigated by
1094 focusing on the top two environmental predictors of genomic variations and using a

1095 Latent Factor Mixed Model analysis. This analysis was implemented with the lfmm R
1096 package[108] on genotype calls from the VQSR pipeline, considering SNPs with call
1097 rate of at least 70% across isolates and minor genotype frequency above 5%, using
1098 K = 3 for the latent factor accounting for underlying population structure. Analyses
1099 were run 5 times and *P* values were combined and adjusted as recommended[108].

1100 ***Gene identification and Gene Ontology enrichment analysis***

1101 Annotation of the revised *H. contortus* genome is ongoing. Therefore, genes
1102 underlying major differentiation signals were inferred by blasting the region of
1103 interest (10 Kbp window for F_{ST} analyses, 2000 bp region around XP-CLR hits)
1104 against the published V1 *H. contortus* genome assembly[13]. Any genes falling within
1105 most probable blast hit coordinates were retrieved from the *H. contortus* published
1106 assembly available from WormBase Parasite[99]. This database was also used to
1107 retrieve *C. elegans* orthologs of positional candidate genes and *H. contortus* Gene
1108 Ontology (GO) terms. GO term enrichment analysis was run with the R topGO[109]
1109 package, considering nodes with at least 10 annotated genes. The "weight01"
1110 algorithm was used to account for existing topology between GO terms. This
1111 framework makes *P*-value of one GO term conditioned on its neighbours, thereby
1112 correcting for multiple testing. Enrichment was tested by the Kolmogorov-Smirnov
1113 statistic applied to gene selection or F_{ST} score accordingly. GO terms with *P* values
1114 below 1% were deemed significant.

1115

1116 # Data availability

1117 Raw sequencing data are archived under the ENA study accession PRJEB9837.
1118 Sequencing data were analysed with publicly available script and software as
1119 mentioned in main text. Outputs were analysed using an R script available at:
1120 https://github.com/guiSalle/Haemonchus_diversity. Reference assembly used in this
1121 project is available at:
1122 ftp://ngs.sanger.ac.uk/production/pathogens/Haemonchus_contortus.

1123 # References

1124 1. Blaxter, M. & Koutsovoulos, G. The evolution of parasitism in Nematoda.
1125 *Parasitology* **142 Suppl 1**, S26-39 (2015).
1126 2. O'Connor, L.J., Walkden-Brown, S.W. & Kahn, L.P. Ecology of the free-living stages
1127 of major trichostrongylid parasites of sheep. *Vet Parasitol* **142**, 1-15 (2006).
1128 3. Jones, J.T. *et al.* Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol*
1129 *Plant Pathol* **14**, 946-61 (2013).
1130 4. Anderson, R.C. The origins of zooparasitic nematodes. *Can J Zool* **62**, 317-328
1131 (1984).
1132 5. Kaplan, R.M. & Vidyashankar, A.N. An inconvenient truth: Global worming and
1133 anthelmintic resistance. *Veterinary Parasitology* **186**, 70-78 (2012).
1134 6. McKellar, Q.A. & Jackson, F. Veterinary anthelmintics: old and new. *Trends*
1135 *Parasitol* **20**, 456-61 (2004).
1136 7. Bundy, D.A.P. *et al.* 100 Years of Mass Deworming Programmes: A Policy
1137 Perspective From the World Bank's Disease Control Priorities Analyses. *Adv*
1138 *Parasitol* **100**, 127-154 (2018).
1139 8. Schulz, J.D., Moser, W., Hurlimann, E. & Keiser, J. Preventive Chemotherapy in the
1140 Fight against Soil-Transmitted Helminthiasis: Achievements and Limitations. *Trends*
1141 *Parasitol* (2018).
1142 9. Hewitson, J.P. & Maizels, R.M. Vaccination against helminth parasite infections.
1143 *Expert Rev Vaccines* **13**, 473-87 (2014).
1144 10. Sallé, G. *et al.* Transcriptomic profiling of nematode parasites surviving vaccine
1145 exposure. *Int J Parasitol* (2018).
1146 11. Gilleard, J.S. Haemonchus contortus as a paradigm and model to study anthelmintic
1147 drug resistance. *Parasitology* **140**, 1506-22 (2013).
1148 12. Doyle, S.R. *et al.* A Genome Resequencing-Based Genetic Map Reveals the
1149 Recombination Landscape of an Outbred Parasitic Nematode in the Presence of
1150 Polyploidy and Polyandry. *Genome Biol Evol* **10**, 396-409 (2018).

1151 13. Laing, R. *et al.* The genome and transcriptome of Haemonchus contortus, a key model
1152 parasite for drug and vaccine discovery. *Genome Biol* **14**, R88 (2013).
1153 14. International Helminth Genomes, C. Comparative genomics of the major parasitic
1154 worms. *Nat Genet* **51**, 163-174 (2019).
1155 15. Mougin, C. *et al.* BRC4Env, a network of Biological Resource Centres for research in
1156 environmental and agricultural sciences. *Environ Sci Pollut Res Int* (2018).
1157 16. Langley, C.H. *et al.* Genomic variation in natural populations of Drosophila
1158 melanogaster. *Genetics* **192**, 533-98 (2012).
1159 17. Small, S.T. *et al.* Population genomics of the filarial nematode parasite Wuchereria
1160 bancrofti from mosquitoes. *Mol Ecol* **25**, 1465-77 (2016).
1161 18. Choi, Y.J. *et al.* Genomic diversity in Onchocerca volvulus and its Wolbachia
1162 endosymbiont. *Nat Microbiol* **2**, 16207 (2016).
1163 19. Perry, G.H. *et al.* Comparative RNA sequencing reveals substantial genetic variation
1164 in endangered primates. *Genome Res* **22**, 602-10 (2012).
1165 20. Denver, D.R. *et al.* A genome-wide view of Caenorhabditis elegans base-substitution
1166 mutation processes. *Proc Natl Acad Sci U S A* **106**, 16310-4 (2009).
1167 21. Saccareau, M. *et al.* Meta-analysis of the parasitic phase traits of Haemonchus
1168 contortus infection in sheep. *Parasit Vectors* **10**, 201 (2017).
1169 22. Geggus, D. The French Slave Trade: An Overview. *The William and Mary Quaterly*
1170 **58**, 119-138 (2001).
1171 23. Anonymous. The wool industry - looking back and forward. (Australian Bureau of
1172 Statistics, 2003).
1173 24. Evesson, B. & Moor, R. The foundation of Australia's fine wool industry. in *Journal*
1174 *of the Royal Australian Historical Society* (2000).
1175 25. Eng, J.K. *et al.* Ivermectin selection on beta-tubulin: evidence in Onchocerca volvulus
1176 and Haemonchus contortus. *Mol Biochem Parasitol* **150**, 229-35 (2006).
1177 26. de Lourdes Mottier, M. & Prichard, R.K. Genetic analysis of a relationship between
1178 macrocyclic lactone and benzimidazole anthelmintic selection on Haemonchus
1179 contortus. *Pharmacogenet Genomics* **18**, 129-40 (2008).
1180 27. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective
1181 sweeps. *Genome Res* **20**, 393-402 (2010).
1182 28. Ardelli, B.F., Stitt, L.E., Tompkins, J.B. & Prichard, R.K. A comparison of the effects
1183 of ivermectin and moxidectin on the nematode Caenorhabditis elegans. *Vet Parasitol*
1184 **165**, 96-108 (2009).
1185 29. Janssen, I.J., Krucken, J., Demeler, J. & von Samson-Himmelstjerna, G.
1186 Transgenically expressed Parascaris P-glycoprotein-11 can modulate ivermectin
1187 susceptibility in Caenorhabditis elegans. *Int J Parasitol Drugs Drug Resist* **5**, 44-7
1188 (2015).
1189 30. Avery, L. & Horvitz, H.R. Effects of starvation and neuroactive drugs on feeding in
1190 Caenorhabditis elegans. *J Exp Zool* **253**, 263-70 (1990).
1191 31. Scott, E.W., Baxter, P. & Armour, J. Fecundity of anthelmintic resistant adult
1192 Haemonchus contortus after exposure to ivermectin or benzimidazoles in vivo. *Res*
1193 *Vet Sci* **50**, 247-9 (1991).
1194 32. Kotze, A.C. & Prichard, R.K. Anthelmintic Resistance in Haemonchus contortus:
1195 History, Mechanisms and Diagnosis. *Adv Parasitol* **93**, 397-428 (2016).
1196 33. Martin, R.J. *et al.* Levamisole receptors: a second awakening. *Trends Parasitol* **28**,
1197 289-96 (2012).
1198 34. Wada, M. *et al.* Isolation and characterization of a GDP/GTP exchange protein
1199 specific for the Rab3 subfamily small G proteins. *J Biol Chem* **272**, 3875-8 (1997).

35. 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).
36. Bolajoko, M.B. *et al.* The basic reproduction quotient (Q0) as a potential spatial predictor of the seasonality of ovine haemonchosis. *Geospat Health* **9**, 333-50 (2015).
37. Fitzpatrick, M.C. & Keller, S.R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol Lett* **18**, 1-16 (2015).
38. Doyle, S.R. *et al.* Population genomic and evolutionary modelling analyses reveal a single major QTL for ivermectin drug resistance in the pathogenic nematode, Haemonchus contortus. *BMC Genomics* **20**, 218 (2019).
39. 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).
40. Troell, K., Engstrom, A., Morrison, D.A., Mattsson, J.G. & Hoglund, J. Global patterns reveal strong population structure in Haemonchus contortus, a nematode parasite of domesticated ruminants. *Int J Parasitol* **36**, 1305-16 (2006).
41. 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).
42. Gilleard, J.S. & Redman, E. Genetic Diversity and Population Structure of Haemonchus contortus. *Adv Parasitol* **93**, 31-68 (2016).
43. Larson, G. *et al.* Current perspectives and the future of domestication studies. *Proc Natl Acad Sci U S A* **111**, 6139-46 (2014).
44. 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).
45. Campbell, M.C. & Tishkoff, S.A. The evolution of human genetic and phenotypic variation in Africa. *Curr Biol* **20**, R166-73 (2010).
46. Chessa, B. *et al.* Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**, 532-6 (2009).
47. Tapio, M. *et al.* Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Mol Biol Evol* **23**, 1776-83 (2006).
48. 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).
49. 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).
50. 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).
51. Roman, A. *Saint Malo au temps des négriers*, 357 (Karthala, 2003).
52. Pereira, F. *et al.* Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. *Mol Biol Evol* **23**, 1420-6 (2006).
53. 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).
54. 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).
55. Laing, R., Gillan, V. & Devaney, E. Ivermectin - Old Drug, New Tricks? *Trends Parasitol* **33**, 463-472 (2017).

1249 56. Osei-Atweneboana, M.Y. *et al.* Phenotypic evidence of emerging ivermectin
1250 resistance in Onchocerca volvulus. *PLoS Negl Trop Dis* **5**, e998 (2011).
1251 57. Doyle, S.R. *et al.* Genome-wide analysis of ivermectin response by Onchocerca
1252 volvulus reveals that genetic drift and soft selective sweeps contribute to loss of drug
1253 sensitivity. *PLoS Negl Trop Dis* **11**, e0005816 (2017).
1254 58. Rose, H. *et al.* Climate-driven changes to the spatio-temporal distribution of the
1255 parasitic nematode, Haemonchus contortus, in sheep in Europe. *Glob Chang Biol* **22**,
1256 1271-85 (2016).
1257 59. van Dijk, J., David, G.P., Baird, G. & Morgan, E.R. Back to the future: developing
1258 hypotheses on the effects of climate change on ovine parasitic gastroenteritis from
1259 historical data. *Vet Parasitol* **158**, 73-84 (2008).
1260 60. Rajpurohit, S., Oliveira, C.C., Etges, W.J. & Gibbs, A.G. Functional genomic and
1261 phenotypic responses to desiccation in natural populations of a desert drosophilid.
1262 *Mol Ecol* **22**, 2698-715 (2013).
1263 61. Sharma, V., Kohli, S. & Brahmachari, V. Correlation between desiccation stress
1264 response and epigenetic modifications of genes in Drosophila melanogaster: An
1265 example of environment-epigenome interaction. *Biochim Biophys Acta* **1860**, 1058-
1266 1068 (2017).
1267 62. Coustham, V. *et al.* Quantitative modulation of polycomb silencing underlies natural
1268 variation in vernalization. *Science* **337**, 584-7 (2012).
1269 63. Cassada, R.C. & Russell, R.L. The dauerlarva, a post-embryonic developmental
1270 variant of the nematode Caenorhabditis elegans. *Dev Biol* **46**, 326-42 (1975).
1271 64. Blitz, N.M. & Gibbs, H.C. Studies on the arrested development of Haemonchus
1272 contortus in sheep. I. The induction of arrested development. *Int J Parasitol* **2**, 5-12
1273 (1972).
1274 65. Wood, D.E. & Salzberg, S.L. Kraken: ultrafast metagenomic sequence classification
1275 using exact alignments. *Genome Biol* **15**, R46 (2014).
1276 66. Li, B.W. *et al.* Transcriptomes and pathways associated with infectivity, survival and
1277 immunogenicity in Brugia malayi L3. *BMC Genomics* **10**, 267 (2009).
1278 67. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-
1279 generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
1280 68. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read
1281 sequencing. *arXiv preprint* **arXiv:1207.3907 [q-bio.GN]**(2012).
1282 69. Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
1283 Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
1284 70. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
1285 analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
1286 71. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
1287 2078-9 (2009).
1288 72. Rensch, T., Villar, D., Horvath, J., Odom, D.T. & Flicek, P. Mitochondrial
1289 heteroplasmy in vertebrates using ChIP-sequencing data. *Genome Biol* **17**, 139
1290 (2016).
1291 73. Howe, D.K., Baer, C.F. & Denver, D.R. High rate of large deletions in Caenorhabditis
1292 briggsae mitochondrial genome mutation processes. *Genome Biol Evol* **2**, 29-38
1293 (2009).
1294 74. Konrad, A. *et al.* Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in
1295 Caenorhabditis elegans Spontaneous Mutation Accumulation Lines of Differing
1296 Population Size. *Mol Biol Evol* **34**, 1319-1334 (2017).
1297 75. Hazkani-Covo, E., Zeller, R.M. & Martin, W. Molecular poltergeists: mitochondrial
1298 DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* **6**, e1000834 (2010).

1299 76. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8
1300 (2011).
1301 77. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal
1302 component analysis of SNP data. *Bioinformatics* **28**, 3326-8 (2012).
1303 78. R Core Team. R: A Language and Environment for Statistical Computing. (R
1304 Foundation for Statistical Computing, Vienna, 2016).
1305 79. Weir, B.S. & Cockerham, C.C. Estimating F-Statistics for the Analysis of Population
1306 Structure. *Evolution* **38**, 1358-1370 (1984).
1307 80. Svardal, H. *et al.* Ancient hybridization and strong adaptation to viruses across
1308 African vervet monkey populations. *Nat Genet* **49**, 1705-1713 (2017).
1309 81. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and
1310 richer datasets. *Gigascience* **4**, 7 (2015).
1311 82. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution
1312 in R language. *Bioinformatics* **20**, 289-90 (2004).
1313 83. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high
1314 throughput. *Nucleic Acids Res* **32**, 1792-7 (2004).
1315 84. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent
1316 and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-
1317 77 (2007).
1318 85. Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J. & McLnerney, J.O.
1319 Assessment of methods for amino acid matrix selection and their use on empirical
1320 data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol*
1321 *Biol* **6**, 29 (2006).
1322 86. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large
1323 phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
1324 87. Skotte, L., Korneliussen, T.S. & Albrechtsen, A. Estimating individual admixture
1325 proportions from next generation sequencing data. *Genetics* **195**, 693-702 (2013).
1326 88. Schiffels, S. & Durbin, R. Inferring human population size and separation history
1327 from multiple genome sequences. *Nat Genet* **46**, 919-25 (2014).
1328 89. Malaspinas, A.S. *et al.* A genomic history of Aboriginal Australia. *Nature* **538**, 207-
1329 214 (2016).
1330 90. McVean, G.A. & Cardin, N.J. Approximating the coalescent with recombination.
1331 *Philos Trans R Soc Lond B Biol Sci* **360**, 1387-93 (2005).
1332 91. Kingman, J.F.C. The coalescent. *Stochastic Processes and their Applications* **13**, 235-
1333 248 (1982).
1334 92. Emery, D.L., Hunt, P.W. & Le Jambre, L.F. Haemonchus contortus: the then and
1335 now, and where to from here? *Int J Parasitol* **46**, 755-769 (2016).
1336 93. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. Inferring
1337 the joint demographic history of multiple populations from multidimensional SNP
1338 frequency data. *PLoS Genet* **5**, e1000695 (2009).
1339 94. Portik, D.M. *et al.* Evaluating mechanisms of diversification in a Guineo-Congolian
1340 tropical forest frog using demographic model selection. *Mol Ecol* **26**, 5245-5263
1341 (2017).
1342 95. Coffman, A.J., Hsieh, P.H., Gravel, S. & Gutenkunst, R.N. Computationally Efficient
1343 Composite Likelihood Statistics for Demographic Inference. *Mol Biol Evol* **33**, 591-3
1344 (2016).
1345 96. Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic
1346 variation. *Bioinformatics* **18**, 337-8 (2002).
1347 97. Suchard, M.A. *et al.* Bayesian phylogenetic and phylodynamic data integration using
1348 BEAST 1.10. *Virus Evol* **4**, vey016 (2018).

98. 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).
99. 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).
100. 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).
101. Korneliussen, T.S., Moltke, I., Albrechtsen, A. & Nielsen, R. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14**, 289 (2013).
102. Hudson, P.J. *et al.* Trophic interactions and population growth rates: describing patterns and identifying mechanisms. *Philos Trans R Soc Lond B Biol Sci* **357**, 1259-71 (2002).
103. Martin, S.H. & Van Belleghem, S.M. Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* **206**, 429-438 (2017).
104. 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).
105. 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).
106. Ellis, N., Smith, S.J. & Pitcher, C.R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156-68 (2012).
107. 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).
108. 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).
109. 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). Authors are grateful to Pr. Beech and Gilleard for insightful discussions.

1388 # Author contributions

1389 GS, JAC designed the experiment. GS, JAC, SD drafted the manuscript. JCa and
1390 JCo sampled and prepared parasite materials. GS performed DNA extraction and
1391 data analyses. SD built the reference genome. MB and NH managed and supervised
1392 parasite sequencing and the project.
1393

1394 # Competing interests

1395 Authors declare they have no competing interests.
1396