# Population genomics reveals an ancient origin of heartworms in canids

Rosemonde I. Power[a], Swaid Abdullah[b], Heather S. Walden[c], Guilherme G. Verocai[d], Tiana L. Sanders[d], Joe L. Luksovsky[d], Andrew R. Moorhead[e], Michael T. Dzimianski[f], Jeremy M. Foster[g], Michelle L. Michalski[h], Alicia Rojas[i,j], Samuel C. Chacón[k], Georgiana Deak[l], Andrei D. Mihalca[l], Patrizia Danesi[m], Elias Papadopoulos[n], Piyanan Taweethavonsawat[o,p], Dung Thi Bui[q,r], Anh Do Ngoc[s], Reuben S.K. Sharma[t], Simon Y.W. Ho[u], Stephen R. Doyle[v,*+] and Jan Šlapeta[a,w,*+]

[a] Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, New South Wales, Australia

[b] University of Queensland, School of Veterinary Science, Gatton, Queensland, Australia

[c] Department of Comparative, Diagnostic, and Population Medicine, College of Veterinary Medicine, University of Florida, Florida, USA

[d] Department of Veterinary Pathobiology, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, Texas, USA

[e] Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina

21 [f] Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia,
22 Athens, Georgia, USA

23 [g] New England Biolabs Inc., Ipswich, Massachusetts, USA

24 [h] University of Wisconsin Oshkosh, Oshkosh, Wisconsin, USA

25 [i] Laboratory of Helminthology, Faculty of Microbiology, University of Costa Rica, San José,
26 Costa Rica

27 [j] Centro de Investigación en Enfermedades Tropicales, University of Costa Rica, San José,
28 Costa Rica

29 [k] Department of Veterinary Parasitology, Veterinary Sciences, Healthy Pet Veterinary Hospital
30 SC, David, Chiriquí, Panamá

31 [l] Department of Parasitology and Parasitic Diseases, University of Agricultural Sciences and
32 Veterinary Medicine of Cluj-Napoca, Calea Mănăștur 3-5, Cluj-Napoca, Romania

33 [m] SCS3 - Parasitology and Mycology Unit, Istituto Zooprofilattico Sperimentale delle Venezie,
34 Viale dell'Università 10, Legnaro, Italy

35 [n] Laboratory of Parasitology and Parasitic Diseases, Faculty of Veterinary Medicine, Aristotle
36 University of Thessaloniki, Thessaloniki, Greece

37 [o] Parasitology Unit, Department of Veterinary Pathology, Faculty of Veterinary Science,
38 Chulalongkorn University, Bangkok, Thailand

39 [p] Biomarkers in Animals Parasitology Research Unit, Chulalongkorn University, Bangkok,
40 Thailand

[q] Department of Parasitology, Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology, Hanoi, Vietnam

[r] Faculty of Ecology and Biological Resources, Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

[s] Department of Medical Parasitology, Vietnam Military Medical University, Hanoi, Vietnam

[t] Faculty of Veterinary Medicine, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[u] School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, New South Wales, Australia

[v] Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK

[w] Sydney Infectious Diseases Institute, The University of Sydney, New South Wales, Australia

+ These authors contributed equally to this work.

* Corresponding authors.

E-mail address: jan.slapeta@sydney.edu.au (J. Šlapeta); stephen.doyle@sanger.ac.uk (S. Doyle)

## Abstract

Heartworms (*Dirofilaria immitis*) are parasitic nematodes that cause significant cardiopulmonary-associated morbidity and mortality in canids worldwide. The global dissemination of heartworms is believed to have occurred alongside the dispersal of domesticated dogs. To test this theory, we performed the largest population genetics study of heartworms to date, based on whole-genome sequencing of 127 modern adult individuals collected from mammalian carnivore hosts across four continents. Population structure and demographic analyses of the nuclear genome reveal distinct genetic differences between heartworms from different continents, indicating a deeper ancient origin and dispersal in canid hosts than previously recognised. Using admixture analyses, we find an Asian origin for Australian heartworms consistent with the arrival of dingoes thousands of years ago. Finally, the genetic relatedness between European and Central American heartworms suggests that modern dispersal, likely associated with human colonisation of the Americas by Europeans, occurred with domesticated dogs. Our findings shed light on the global population dynamics and evolutionary history of heartworms, which can aid future surveillance and control efforts for this important veterinary parasite.

72 ## Introduction

73 The heartworm *Dirofilaria immitis* is a parasitic nematode that infects a broad range of
74 mammalian carnivores (order Carnivora) across tropical, subtropical and temperate regions
75 worldwide [1]. Mosquitoes transmit larval stages of heartworms, which develop after infection
76 and reside in their host's heart and pulmonary arteries as adults [2]. These adult worms can cause
77 lesions, vascular damage, pulmonary hypertension, and right-sided congestive heart failure,
78 potentially leading to the death of their host [2]. This is of particular concern for domestic dogs,
79 whereby potentially life-threatening cardiopulmonary disease is globally managed by lifelong
80 and rigorous application of parasiticides [3]. Such treatment regimes have led to the emergence
81 of drug resistance [4,5] that, while currently only well described in the USA, threatens heartworm
82 control globally if it were to spread. Heartworms are predicted to increase in prevalence and
83 expand into new regions due to climate change, pet travel, and habitat alterations [6]. Therefore,
84 defining the origins, distribution, and capacity for transmission will have important
85 implications for implementing effective heartworm surveillance and control.

86 Current views on the evolutionary origins and subsequent radiation of heartworms
87 worldwide are inherently anthropocentric. Heartworms are believed to have originated in
88 Europe [7] or Asia [2] and spread globally over the past 200–400 years via human-mediated dog
89 movements. As a result of these recent dispersals, heartworms are presumed to form a
90 homogeneous global population. However, these views lack robust empirical support and
91 overlook the deep co-evolutionary history that some parasites have shared with their hosts and
92 environments. The broad host range of heartworm potentially indicates a long co-evolutionary
93 history with mammalian carnivores, especially canids [8,9], a group that dates back approximately
94 40 Mya [10], predating humans (~2.8 Mya) [11] and the domestication of modern dogs (14–40 kya)
95 [12–14]. Given also that mosquito vectors have existed since the early Triassic (217 Mya) [15], it is
96 possible that heartworms originated and spread long before human involvement. We propose

that heartworms offer a compelling case for assessing the degree to which anthropogenic activities have influenced parasite distribution, with a potentially much deeper evolutionary history than previously recognised.

Analyses of genomic data have significant potential to identify the historical origins of heartworms and infer their global dispersal routes, as demonstrated for other helminths [16–18]. Sequencing adult heartworms is recommended for obtaining high-quality data [19]; however, obtaining adults from the heart requires invasive surgery or post-mortem, which are often not feasible. Such sampling is especially difficult in parts of Asia and Africa, where cultural perceptions of dogs as unclean or carriers of disease can restrict sampling opportunities. In light of these challenges, here we present the most extensive population genomics study of heartworms to date. Using whole genome sequencing of 127 adult worms obtained from nine countries on four continents, we analyse the diversity, genetic structure, and population demographics of modern heartworms to infer their historical origins and dispersal. Our analyses suggest that heartworm diversity has been shaped by the distribution of ancestral canid hosts throughout time and provide evidence of recent human-driven migration.

113 ## Results and Discussion

114 **Genomic diversity of heartworms from multiple continents, countries, and hosts**

115 We generated whole-genome sequencing data for 127 adult heartworms (*Dirofilaria*
116 *immitis*) sampled from nine countries in five regions: Australia, North America (USA), Central
117 America (Costa Rica, Panama), Europe (Greece, Italy, Romania), and Asia (Malaysia,
118 Thailand) (Fig. 1a). Samples were primarily collected from domestic dogs ($n$ = 115), but we
119 also sampled cats ($n$ = 4), foxes ($n$ = 4), a ferret ($n$ = 1), a golden jackal ($n$ = 1), a leopard ($n$ =
120 1), and a wildcat ($n$ = 1). Overall, we obtained an average coverage per sample of 75.47× for
121 the nuclear genome, 13,119.80× for the mitochondrial genome, and 1,047.09× for the
122 *Wolbachia* endosymbiont genome (Supplementary Table 1, Supplementary Fig. 1–4). Variant
123 calling, joint genotyping, and filtering of heartworm samples revealed 264,558 nuclear single
124 nucleotide polymorphisms (SNPs) (3.3 SNPs/kb), 32 mitochondrial SNPs (2.3 SNPs/kb), and
125 280 *Wolbachia* SNPs (0.3 SNPs/kb). The SNP frequencies in the nuclear and mitochondrial
126 genomes are comparable, whereas the frequency in the *Wolbachia* genome is considerably
127 lower. These frequency data are comparable to those of other *Wolbachia*-infected filarial
128 worms, such as *Onchocerca volvulus* [20] and *Wuchereria bancrofti* [18].

129

130 **Nuclear variants reveal distinct continental partitioning of heartworms**

131 Analyses of the heartworm cohort's genetic diversity revealed distinct groups of
132 samples based on broad geographical regions. Principal component analysis (PCA) using
133 196,550 high-quality autosomal SNPs revealed four clusters of samples corresponding to the
134 continents from which the samples were collected (Fig. 1b and Extended Data Fig. 1; PC1
135 variance: 12.6%, PC2 variance: 10.69%). The third and fourth PCs showed a similar pattern of
136 geographical clustering (Extended Data Fig. 2a; PC3 variance: 5.76%, PC4 variance: 4.39%);

137 however, although the signal was comparatively weaker, the European and Central American
138 samples formed separate clusters. Similar findings were observed in a previous study of 31
139 heartworm samples from Australia, Italy, and the USA, where samples from each country
140 formed distinct clusters [21]. Independent PCAs of SNPs from each chromosome found the same
141 continental clustering pattern across all chromosomes, including the sex-linked X chromosome
142 (Extended Data Fig. 2b–f), suggesting that this is a genome-wide rather than region-specific
143 observation. The distinct geographical partitioning in our nuclear data was not supported by
144 the mitochondrial (Extended Data Fig. 3a) or *Wolbachia* data (Extended Data Fig. 3b), likely
145 due to the few variant sites within each dataset.

146 Analyses of genome-wide nucleotide diversity (Pi) further demonstrated differences
147 between heartworm populations across continents. Pi differed significantly between all
148 populations (Wilcoxon rank sum test, $p < 0.001$), with the exception of Europe and Central
149 America (Fig. 1c). The highest Pi was observed in the USA (Pi[mean] = $7.39\times10^{-4}$), while the
150 lowest was in Asia (Pi[mean] = $2.89\times10^{-4}$). The genetic diversity of our USA samples aligned
151 closely with that of a previous *D. immitis* study (Pi[mean] = $7.36\times10^{-4}$) [21]. However, our
152 Australian samples (Pi[mean] = $5.61\times10^{-4}$) exhibited higher genetic diversity than those
153 reported previously (Pi[mean] = $3.9\times10^{-5}$) [21], likely due to the larger and geographically
154 broader sample here. Overall, the nucleotide diversity of our *D. immitis* samples is comparable
155 to that of the filarial worm *W. bancrofti* (Pi[mean] = $2.4\times10^{-4}$) [18] but lower than that of *O.*
156 *volvulus* (Pi[mean] = $4.0\times10^{-3}$) [20]. Each *D. immitis* population exhibited different rates of
157 linkage disequilibrium (LD) decay and baseline LD (Fig. 1d), with Asia and the USA showing
158 the highest and lowest LD, respectively. Estimates of admixture provided further support for
159 continental partitioning in heartworms. To infer the ancestry proportions of samples, we used
160 NGSadmix with a range of *K* clusters (*K* = 2–10; Extended Data Fig. 4). At *K* = 6, Australian
161 samples showed the most diverse admixture patterns with strong evidence of shared ancestry

with Asia, whereas USA, Europe, and CENAM were largely distinct with some evidence of shared ancestry in a minority of European and CENAM samples (Fig. 1e). The strong population structure in the nuclear genome of heartworms suggests that spatial spread has been restricted during the species' history, resulting in distinct genetic profiles on each continent.

**Inference of transmission between diverse heartworm hosts**

Although most heartworm samples were obtained from dogs, the analysis of heartworm genomes from foxes, cats, a golden jackal, a ferret, and a leopard allowed us to evaluate genetic variation among different host species. We found that geography rather than the host species explained variation in the global heartworm population, evidenced by samples from cat and ferret nested within samples from dogs in the USA, fox and dogs in both Australian and Italian groups, and dogs and jackal, wildcat and leopard from Europe (Fig. 2a). Analyses of nucleotide diversity (Pi), absolute nucleotide divergence ($D_{XY}$), and genetic differentiation ($F_{ST}$) provided additional insights on the variation within and between different host species. The ferret differed from other hosts as its heartworm had the highest diversity (Pi[mean] = $1.21\times10^{-3}$) (Fig. 2b) and the greatest divergence from other hosts (Extended Data Fig. 5a). In contrast, the samples from the golden jackal, leopard, and wildcat had the lowest diversity (Pi[mean] 1.10–$1.89\times10^{-4}$) and were the least divergent from each other compared with all other host pairs. These hosts, however, were each represented by only a single worm, and the variation observed in the jackal, leopard, and wildcat samples may be confounded as they all originated from Romania. There was little genetic differentiation between heartworms from dogs and ferrets, and between dogs and foxes, suggesting that heartworms from canids and mustelids may be more genetically similar to each other than to those from felids (Extended Data Fig. 5b). Collectively, these data suggest an absence of obvious host-specificity and that transmission is

186 likely between domestic animals and wildlife. However, further geographic sampling from
187 non-dog hosts is needed to exclude the possibility of host-specific adaptation.

188

189 **Distinguishing modern from ancient origins and dispersal**

190 The current hypothesis for heartworm distribution around the world is that it has been
191 largely influenced by human movement in contemporary times. Alternative hypotheses are that
192 heartworms spread with humans during or shortly after dog domestication ~14–40 kya, or that
193 heartworms spread with wild canids prior to domestication. To further explore the demographic
194 history of heartworms, we inferred their effective population size histories and population split
195 times using SMC++. Effective population sizes were smaller than estimates of past population
196 sizes for all continents (Fig. 3a). Earlier population splits in the demographic analysis broadly
197 supported the $D_{XY}$ and $F_{ST}$ results (Extended Data Fig. 6). Notably, the earliest estimated
198 population splits occurred in Asia, which split from the populations from Europe, the USA,
199 and Central America between ~46 and 63 kya. Importantly, these data suggest that the
200 divergence between populations is older than the proposed dates for dog domestication [12–14].

201 Considering a pre-domestication dispersal hypothesis, we next questioned which host
202 may have harboured and dispersed heartworms before the emergence of modern dogs. We
203 hypothesised that longer co-evolution between parasites and hosts would likely result in the
204 latter being able to carry higher parasite loads. Felids (e.g. domestic cats, wild cats, snow
205 leopards), procyonids (e.g. raccoon dogs), mustelids (e.g. European badger, ferrets), pinnipeds
206 (e.g. harbour seal, brown fur seal), and ursids (e.g. brown bear, black bear) have been described
207 as being infected with heartworms (Supplementary Table 2); however, they typically carry
208 fewer adult worms and fail to support the parasite's entire life cycle, and so they are not
209 considered primary reservoirs [22–25] (Fig. 3b). In contrast, wild canids are commonly

210 microfilaremic and have high worm burdens. Wolves and coyotes have worm burdens similar
211 to dogs, while jackals and foxes have significantly lower burdens than dogs (Kruskal-Wallis
212 test with Dunn's multiple comparisons; Dog vs Jackal: $P \leq 0.05$, Dog vs Fox: $P \leq 0.001$) (Fig.
213 3b). These findings imply that canids, particularly ancestral wolves or coyotes, were most
214 likely the primary hosts for heartworm during ancient times [26].

215

216 **Coevolution of canids and dogs from ancient to contemporary times**

217 Previous work had proposed that canids originated in North America ~40 Mya when
218 the continent was completely isolated [10], whereas mosquitoes originated much earlier in the
219 Triassic (~217 Mya) [15]. Phylogenetic analyses of mitochondrial DNA suggest that heartworms
220 diverged from their closest relatives in the genus *Onchocerca* ~25 Mya during the Oligocene
221 [27], i.e., when canids and mosquito vectors were already well-established. Together, this
222 timeline supports a long co-evolutionary history between heartworms and ancient canids.

223 The formation of the Bering Land Bridge and the Isthmus of Panama during the
224 Miocene and Pliocene facilitated the first migrations of canids into Eurasia (~7 Mya) [10] and
225 South America (3.5–3.9 Mya) [28], respectively. Throughout history, fluctuations in sea level
226 have caused the Bering Land Bridge to be repeatedly exposed and submerged. Fossil records
227 and genomic data provide robust evidence that these periods of exposure allowed ancient
228 canids to form continuous populations across the Northern Hemisphere. Examples include
229 Pleistocene wolves, which were highly connected across North America and Eurasia [12,29,30],
230 *Xenocyon lycaonoides,* a large canid distributed across North America and Eurasia in the mid
231 Pleistocene [31] and *Eucyon,* a jackal-sized canid distributed across North America, Eurasia, and
232 Africa in the late Miocene [32]. The existence of these interconnected ancient canid populations

233 suggests that heartworms may have formed a highly connected population with their hosts at
234 that time.

235 If heartworms were once one population, when did they diverge, and why? We
236 estimated the timing of this contiguous *D. immitis* population to predate ~63 kya before the
237 species diverged on each continent (Fig. 3a). This roughly aligns with the last interglacial
238 period (116 – 130 kya), when Earth experienced a climate as warm or even warmer than today
239 [33]. Such an environment could have been suitable for the development and transmission of
240 heartworms via mosquito vectors. Following the warm climates of the last interglacial period,
241 the Earth began to cool. By Marine Isotope Stage 4 (58–72 kya), large continental ice sheets
242 had formed across North America, Beringia, and northern Eurasia [34]. These ice sheets separated
243 ancient canid populations [29,35], potentially fragmenting the associated ancient heartworm
244 population. This fragmentation would have reduced gene flow for extended periods, explaining
245 our study's continental grouping of samples. Further insights might be obtained by exploring
246 the presence of heartworm DNA in the remains of ancient canids; if successfully extracted,
247 these ancient worm DNA sequences may fill in the temporal and geographic gaps needed to
248 understand heartworm divergence in more detail. Similarly, further sampling in unstudied
249 regions including South America and Africa, and finer sampling within Asia, would enhance
250 our understanding of heartworms' evolutionary history.

251 Considering global warming and cooling cycles, the fragmentation of canid and
252 heartworm populations would likely have occurred more than once. We found evidence of
253 secondary admixture between heartworms from Asia and the USA after the initial population
254 divergence, evident from both the admixture analyses and gene flow between populations (Fig.
255 4a; delta m = 3.4; all tree migration edges are presented in Extended Data Fig. 7). *D*-statistics
256 (Fig. 4b,c) provided further support for excessive allele sharing between the USA and Asia
257 relative to that between the USA and other regions (Fig. 4b). Given that heartworms from the

258 USA and Asia diverged >45 kya (Fig. 3a), this admixture may reflect multiple waves of
259 migration with canids across the Bering Land Bridge in the Late Pleistocene, during periods of
260 ice sheet contraction [12].

261

262 **The origin and spread of heartworms in Australia**

263 Heartworms are endemic in Australia, but the mechanisms and timing of their arrival
264 in the country have remained an enigma. Consistent with the modern human dispersal
265 hypothesis, one means of introduction could have been after European colonisation that began
266 ~235 years ago. However, considering our genetic evidence of older, canid-mediated dispersal,
267 a more likely scenario is that heartworms were brought to Australia via dingoes. Dingoes are
268 an ancient dog breed from Asia and were the only canids present on the Australian continent
269 until European settlement. Recent morphometric analysis indicates an East Asian origin of
270 dingoes via Melanesia [36], whereas genetic studies suggest that dingoes likely came from Island
271 Southeast Asia and were brought to Australia by humans between 5 and 12 kya [37].

272 Multiple lines of evidence support the introduction of heartworms into Australia via
273 Asia. Both admixture and TreeMix analyses highlighted the close relationship, evidenced by
274 partial or complete admixture profiles (Fig. 1e) and mutual monophyly (Fig. 4a) between Asian
275 samples and samples collected along the East Coast of Australia (Fig. 5a). Calculations of $D_{XY}$
276 showed that AUS-ASIA ($D_{XY}$[mean] = 5.13e–4) was the second least divergent pair (after
277 EUR-CENAM, which indicates much more recent dispersal pattern [see below]), suggesting
278 that Australian worms are more closely related to those in Asia than elsewhere (Extended Data
279 Fig. 6a). *D*-statistics solidified this relationship by showing excessive allele sharing between
280 the two continents relative to that between Australia and other regions (Fig. 5b). These findings

281 point towards an Asian origin of Australian heartworms, likely transported with a dingo host,
282 rather than by an introduction through European colonisation.

283 We estimated the divergence of Asian and Australian heartworms to be ~29 kya (Fig.
284 3a), which is older than the previous estimates of 5 and 12 kya for the dingo arrival. This
285 discrepancy likely arises from the limited geographic scope of our Asian samples, which are
286 only from Thailand and Peninsular Malaysia. Sampling across a broader geographic region,
287 particularly throughout the Indonesian archipelago and Melanesia, could bridge the gap
288 between our Australian and Asian samples. Such data may provide further granularity of the
289 Asia–Australia divergence and potentially coincide more closely with the timeline of dingo
290 arrival.

291 Considering the extensive sampling along the north–south transect of Eastern Australia,
292 we sought to identify factors that might explain current genetic diversity. Australian
293 heartworms were the second most genetically diverse population globally (Fig. 1c). Within
294 Australia, diversity was highest in Lockhart River (Pi[mean] = 8.12e-4) and lowest in
295 Rockhampton (Pi[mean] = 2.15e-4) (Fig. 5c). Heartworms from the northern state of
296 Queensland were highly admixed, with two unique sources of ancestry not found on any other
297 continent (Fig. 1e). Interestingly, ten of 16 heartworms from Sydney, the southernmost site,
298 had a single fixed ancestry that was detected in Queensland samples. Another four Sydney
299 heartworms had a different fixed ancestry identical to that found in Asia, potentially reflecting
300 modern movement, though the exact cause remains unclear. Previously, it was believed that
301 heartworm cases in temperate Sydney and broader New South Wales were primarily caused by
302 infected pets or mosquitoes travelling down from tropical and subtropical Queensland [38], where
303 climates are more favourable to heartworm transmission. However, our nuclear PCA showed
304 a large cluster of samples primarily from Queensland, alongside two distinct clusters from
305 Sydney (Fig. 5d). This suggests that Sydney heartworms may not necessarily originate from

306 Queensland, as previously believed. Further investigations are required to determine whether
307 there are local endemic populations of heartworms in Sydney. Differences in $D_{XY}$ and $F_{ST}$
308 between Australian heartworm subpopulations were subtle, showing no clear relationship
309 between genetic and geographical distance as might be expected for a vector-borne parasite
310 (Fig. 5e, f).

311

312 **Modern heartworm migration between Europe and Central America**

313 Our genomic analysis revealed a close genetic relationship between worms in Europe
314 and Central America. This finding was most evident in our PCA of nuclear SNPs, where
315 samples from Greece, Italy, and Romania clustered with those from Costa Rica and Panama
316 (Fig. 1b). In addition, estimates of $D_{XY}$ revealed that EUR-CENAM ($D_{XY}$[mean] = 4.95e–4)
317 was the least divergent compared with other populations. Admixture analysis using NGSadmix
318 further supported this trend as the European and Central American samples shared high levels
319 of ancestry at lower *K* values (*K* = 2, 3, and 5; Extended Data Fig. 4). Although increasing *K*
320 helped to differentiate worms from these two continents, low levels of admixture were still
321 present (Fig. 1e). These findings were corroborated using *D*-statistics which provided strong
322 evidence of excessive allele sharing between heartworms from Central America and Europe
323 relative to that between Central America and other regions (Fig. 4c). These findings support a
324 modern migration event between European and Central American heartworms which took
325 place after dog domestication.

326 One plausible modern migration event between Europe and Central America was the
327 transatlantic transportation of domestic animals, including dogs, during the European
328 colonisation of the Americas beginning in the late 15th century. Animals introduced to the
329 Americas during the Age of Discovery were primarily livestock species, including cattle,

330 sheep, pigs, goats, horses, and mules, but dogs such as greyhounds and mastiffs were also
331 included for hunting, guarding, protection, shepherding, and battle purposes [39–41]. The
332 introduction of European dogs into the Americas is perhaps best documented from Columbus'
333 voyages, with historical records indicating the presence of at least 20 dogs on board during the
334 second voyage to Hispaniola [39]. Therefore, infected dogs from Europe may have transported
335 local strains of this parasite to the Americas. Alternatively, the shared ancestry between
336 European and Central American heartworms could result from the more recent migration of
337 humans and their pets. Given the population split times proposed in our demographic analyses
338 and the lack of evidence for recent migration and transmission between continents elsewhere,
339 we posit that the former scenario is more biologically plausible and likely occurred over the
340 past few centuries.

341

342

343 **Concluding remarks**

344 Our transcontinental genome-wide analysis of heartworm parasites has revealed a much
345 deeper origin and dispersal history than previously understood, allowing us to propose new
346 hypotheses on the evolutionary history of heartworms (Fig. 6). Rather than being solely driven
347 by human-mediated dispersal of dogs over the past few hundred years, we found that ancestral
348 canid hosts played a pivotal role in the evolution and dissemination of heartworms around the
349 world. Heartworms have evolved to exploit a unique niche in the heart and bloodstream of
350 canids, an environment that has likely remained unchanged over millions of years between
351 hosts. In the absence of niche competition, heartworms could be a relic of the past, which, until
352 recently, had little need for further adaptation. Anthropogenic influence is, however, likely to
353 shape heartworm diversity more rapidly than ever before. For example, extensive drug

treatment is already associated with resistance, particularly in the USA; climate change is expected to shift populations of parasites and their vectors into higher latitudes; and increasing global connectivity of people and their pets will all influence this distribution and adaptation of heartworms. Therefore, understanding the global genomic landscape of these parasites, as demonstrated here, is vital for developing effective surveillance and control strategies in the future. Our data indicate that such surveillance and control strategies should account for geographical variation between parasite populations and be aware of transmission within and between populations to maximise sustainable parasite control.

362 # Methods

363 ## Sample collection

364 A total of 120 individual adult *Dirofilaria immitis* were obtained from Australia (*n* =
365 43), the USA (*n* = 37), Panama (*n* = 14), Greece (*n* = 12), Thailand (*n* = 6), Italy (*n* = 3),
366 Romania (*n* = 3), Costa Rica (*n* = 1), and Malaysia (*n* = 1). The sampling regime aimed to
367 collect adult *D. immitis* from a wide geographical range. Most of these samples were obtained
368 from dogs (*Canis lupus familiaris*; *n* = 115) but also included some from cats (*Felis catus*; *n* =
369 4), foxes (*Vulpes vulpes*; *n* = 4), a ferret (*Mustela putorius furo*; *n* = 1), a golden jackal (*Canis*
370 *aureus*; *n* = 1), a leopard (*Panthera pardus*; *n* = 1), and a wildcat (*Felis silvestris*; *n* = 1)
371 (metadata for all samples is presented in Supplementary Table 3). We included five technical
372 replicates (Brisbane: *n* = 1; Sydney: *n* = 2; Malaysia: *n* = 1; Panama: *n* = 1). To provide
373 outgroup samples for our analyses, we collected adult *D. repens* (*n* = 2) from dogs in Italy,
374 *Dirofilaria* sp. 'Thailand II' (*n* = 2) from humans (*Homo sapiens*) in Vietnam and *D. ursi* (*n* =
375 1) from an American black bear (*Ursus americanus*) in the USA.

376

377 ## DNA extraction and whole genome sequencing

378 *Dirofilaria* specimens were frozen at −20°C or stored in ethanol upon collection and
379 sent to the University of Sydney, Australia. A 1–2 cm segment between the first and second
380 third of the anterior portion of the worms was excised using a sterile scalpel blade and placed
381 in a 1.5 mL Eppendorf tube.

382 ***Ethanol-preserved samples.*** Worm segments were centrifuged at maximum speed for
383 30 s using a benchtop centrifuge (Eppendorf, Australia) and excess ethanol was removed. The
384 segments were washed in 500 mL of phosphate-buffered saline (PBS) and placed on a heat

385 block (Heat Block Eppendorf Thermomixer Comfort or Heat Block Major Science Dry Bath
386 Incubator) at 60°C for 20–40 min, depending on their thickness, to evaporate any remaining
387 ethanol.

388 ***Frozen samples.*** Worm segments were defrosted and washed three times in PBS.
389 Genomic DNA was extracted from the tissue segments using a Monarch® Genomic DNA
390 Purification Kit (New England Biolabs, Australia) as per the manufacturer's protocol. Extracted
391 DNA was shipped to Novogene (HK) Co., Ltd for indexing, library preparation, and whole
392 genome sequencing using the NovaSeq 6000 platform. Overall, 128/130 and 2/130 DNA
393 samples were sequenced at expected depths of ~10 G and ~1 G, respectively.

394

395 **Bioinformatic analyses**

396 The bioinformatics workflow of this study is summarised in Extended Data Fig. 8. In
397 addition to the data generated in this study, we included publicly available datasets of
398 individual *D. immitis*. These data were derived from (i) five adults from Sydney, Australia
399 (SRA NCBI: SRR13154013-7) [42]; (ii) one adult from Georgia, USA (ENA Accession:
400 ERR034940) and one adult from Pavia, Italy (ENA Accession: ERR034941) [43].

401

402 **Raw data processing and mapping**

403 Multiple FastQ files were generated for some samples and merged before data analysis.
404 The quality of all raw paired-end sequencing reads was inspected using FastQC (version 0.11.8;
405 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the output was summarised
406 using MultiQC (version 1.17 [44]). The raw paired-end reads were scanned with a 10-base sliding
407 window and trimmed to have a minimum average quality score of 20 per base and a minimum

408 read length of 50 bp using Trimmomatic [45] (version 0.39). We used a Nextflow [46] mapping
409 pipeline (mapping-helminth/v1.0.9 [47]) to map the trimmed paired-end reads to a combined
410 reference genome containing *D. immitis* (dimmitis_WSI_2.2 [21]) and the domestic dog
411 (GenBank accession: GCA_014441545). The dimmitis_WSI_2.2 genome has been previously
412 supplemented with the *D. immitis* mitochondrial and *D. immitis*-associated *Wolbachia*
413 endosymbiont genomes [21]. The mapping pipeline was designed for the standard mapping of
414 helminth genomes using Minimap2 (version 2.17 [48]) as a mapper and GATK (version 4.1.4.1
415 [49]), Samtools (version 1.14 [50]), and Sambamba (version 1.0 [51]) for further data processing.
416 Reads aligning to only the *D. immitis* genome were extracted for downstream analysis using
417 Samtools.

418

419 **Sequencing Coverage**

420 Sequencing coverage was examined using BamTools (version 2.5.1 [52]), BEDTools
421 makewindows (version 2.31.0 [53]), and SAMtools bedcov. Coverage statistics were obtained for
422 the nuclear, mitochondrial, and *Wolbachia* genomes of *D. immitis*. To visualise nuclear *D.*
423 *immitis* sequencing coverage for each sample, the scaffold, mitochondrial and *Wolbachia*
424 genomes of *D. immitis* were removed, along with any chromosomes of the domestic dog
425 genome. The ratio of X chromosome to autosome coverage was visualised to determine the sex
426 of the *D. immitis* samples, identified by a drop in coverage along the X chromosome (ratio ~
427 0.5) indicative of a male parasite [21].

428

429 **Variant calling**

430 Variants were identified using GATK (version 4.1.4.1; [49]) HaplotypeCaller,
431 CombineGVCFs, and GenotypeGVCFs. GVCF files were first produced for each sample and

432 then combined into a cohort-level GVCF file, which was then used for joint genotyping.
433 Outgroup samples were removed to obtain a *D. immitis* cohort VCF file. Variants in the
434 nuclear, mitochondrial, and *Wolbachia* datasets were filtered separately using GATK
435 SelectVariants, GATK VariantFiltration, and VCFtools (version 1.16 [54]). The distribution of
436 eight quality metrics was established (QUAL, DP, MQ, SOR, QD, FS, MQRankSum, and
437 ReadPosRankSum), and the relevant upper and/or lower 1% distribution tails were removed.
438 The single nucleotide variants (SNVs) were further filtered based on the following parameters:
439 bi-allelic, depth ≥3, Hardy-Weinberg Equilibrium p-value $\geq 10^{-6}$ (nuclear), and minor allele
440 frequency ≥0.02. Indels were not included in subsequent analyses. Samples with <50% of the
441 total variants were removed, and an optimal threshold for missing data of 0.9 was established
442 (nuclear, mitochondrial, and *Wolbachia*). Additional VCF files were generated from the
443 filtered nuclear data for autosomes (chromosomes 1–4), the sex-linked X chromosome, and
444 Australian samples. The same filtering parameters were applied to the VCF file containing both
445 *D. immitis* and outgroup samples. SNP discordance in the filtered nuclear autosome data for
446 the technical replicate samples was measured using VCFtools --diff-indv-discordance. All
447 technical replicates had < 0.5% SNP discordance with their respective original samples.

448

449 **Population structure**

450 To explore the broad-scale genetic relatedness between *D. immitis* individuals and
451 populations, we performed principal component analysis (PCA) using the R package
452 SNPRelate (version 1.38.0; [55]). PCA was conducted for the nuclear (autosomes together,
453 autosomes separate, and chromosome X separate), mitochondrial, and *Wolbachia* data.

454 Nucleotide diversity (Pi) within populations, as well as absolute nucleotide divergence
455 ($D_{XY}$) and genetic differentiation ($F_{ST}$) between populations, were calculated using pixy

456 (version 1.2.10.beta2; [56]). First, the GVCF files were genotyped using GATK GenotypeGVCFs
457 with the inclusion of invariant sites (--all-sites). The nuclear invariants were selected using
458 GATK SelectVariants (--select-type-to-include NO_VARIATION) and merged with the
459 previously filtered nuclear SNPs. The VCF file was filtered to only include invariants and SNPs
460 in the autosomes and sex-linked X chromosome. We conducted the pixy analysis with a
461 window size of 100 kb. Separate runs were performed, grouping the samples in various ways:
462 by region (using dog host samples only), by host species (using all samples), and by city/state
463 (using Australian dog host samples only).

464 Linkage disequilibrium (LD) statistics were obtained from the nuclear data using
465 VCFtools. First, we filtered and thinned the data (--maf 0.02 --thin 1000 --max-missing 1). We
466 then calculated the squared correlation coefficient between genotypes (--ld-window-bp
467 1000000 --max-alleles 2 --min-alleles 2 --min-r2 0.1 --geno-r2). To visualise patterns of LD
468 decay, the output was plotted per 0.1 kb for each population.

469

470 **Admixture**

471 The global ancestry and admixture proportions of *D. immitis* were inferred using
472 NGSadmix from the ANGSD package [57]. First, genotype likelihoods were extracted from the
473 filtered autosomal *D. immitis* VCF file using VCFtools (--max-missing 1 --BEAGLE-PL).
474 Genotype likelihoods were extracted separately for each chromosome and merged into a single
475 dataset. NGSadmix (-minMaf 0.02 -misTol 0.9) was performed using multiple combinations
476 of different clusters ($K$ = 2 to 10) and seeds (1 to 5). CLUMPAK was used to determine the
477 optimal $K$ value [58].

478 We used TreeMix (version 1.12; [59]) to produce a maximum-likelihood tree and infer
479 population splits and admixture events in the *D. immitis* dataset. To generate the input data,

480 VCFtools was used to filter the cohort SNV VCF file (--max-missing 1), retaining only
481 autosomal data and the same samples as in the previous variant-calling filtration step. The
482 variants were pruned for linkage disequilibrium (10 kb windows, shifting by 10 kb, LD
483 threshold 0.1), and the VCF file was converted to the appropriate format using custom scripts
484 ('ldPruning.sh' and 'vcf2treemix.sh') obtained from
485 https://github.com/speciationgenomics/scripts. TreeMix was run across a range of migration
486 edges (0 to 5) and seeds (0 to 10), with *D. ursi* included as an outgroup. The optimal number
487 of migration edges was estimated using the R package OptM (version 1.8;
488 https://github.com/cran/OptM).

489 To further investigate admixture in the cohort, *D*-statistics were inferred using
490 AdmixTools (version 5.1 [60]). First, we converted our filtered nuclear VCF file into the required
491 eigenstrat format using a custom script ('convertVCFtoEigenstrat.sh') obtained from
492 https://github.com/speciationgenomics/scripts/. Additional tools required for file conversion
493 were VCFtools and PLINK (version 1.90b6.18 [61]). qpDstat was used to perform *D*-statistic
494 testing for all population pairs relative to the other populations (PopX), with *D. ursi* as the
495 outgroup taxa. *Z*-scores with an absolute value >3 were considered significant.

496

497 **Population demography**

498 We estimated the population size history of *D. immitis* and divergence times between
499 subpopulations using SMC++[71] (version 1.15.3 [62]). Samples from dog hosts in the filtered
500 autosomal VCF were grouped by continent (*i.e.*, Asia, Australia, Central America, Europe, and
501 USA) and analysed separately before being plotted together. First, the VCF file was converted
502 to the SMC++ input format using *vcf2smc*. The model was fitted using *estimate* with a
503 timepoint range of 1 to 5,000,000 and the nematode *Caenorhabditis elegans* mutation rate of

504 $2.7\times10^{-9}$ per site per generation. The *plot* command (-g -c) was used to visualise the output,
505 with the -g parameter specifying the generation time of the target population. *D. immitis* takes
506 ~1 year to produce offspring and can live for up to 7 years in the host's cardiopulmonary
507 system. The development and transmission of this parasite are also highly dependent on local
508 climate and vector availability. To account for potential variability in the generation time of *D.*
509 *immitis*, we tested and compared multiple generation times (-g 1, 2, 4, or 6) (Extended Data
510 Fig. 9). Given the reliance on viable mosquito vectors for transmission, and the presence of
511 both male and female adult worms for sexual reproduction, we determined that a 4-year interval
512 was the most appropriate estimate. In addition, we used *split* to estimate divergence times
513 between all pairs of subpopulations.

514

515 **Data availability**

516 Raw FastQ sequences generated in this study are available at SRA NCBI BioProject
517 PRJNA1104412.

518

519 **Code availability**

520 The code used to analyse the data in this study is available from GitHub:
521 https://github.com/rosemondepower/heartworm_genome.git.

522 ## References


523 1. Noack, S., Harrington, J., Carithers, D. S., Kaminsky, R. & Selzer, P. M. Heartworm
524 disease - Overview, intervention, and industry perspective. *Int. J. Parasitol. Drugs Drug*
525 *Resist.* **16**, 65–89 (2021).
526 2. Bowman, D. D. & Atkins, C. E. Heartworm biology, treatment, and control. *Vet. Clin.*
527 *North Am. Small Anim. Pract.* **39**, 1127–58, vii (2009).
528 3. Nelson, T., McCall, J. W., Moorhead, A., Starkey, L. & Ames, M. *American Heartworm*
529 *Society Canine Guidelines for the Prevention, Diagnosis, and Management of*
530 *Heartworm (Dirofilaria Immitis) Infection in Dogs*.
531 https://d3ft8sckhnqim2.cloudfront.net/images/AHS_Canine_Guidelinesweb22NOV2024
532 .pdf?1732318144 (2024).
533 4. Hampshire, V. A. Evaluation of efficacy of heartworm preventive products at the FDA.
534 *Vet Parasitol* **133**, 191–195 (2005).
535 5. Pulaski, C. N. *et al.* Establishment of macrocyclic lactone resistant *Dirofilaria immitis*
536 isolates in experimentally infected laboratory dogs. *Parasit. Vectors* **7**, 494 (2014).
537 6. Simón, F. *et al.* Human and animal dirofilariasis: the emergence of a zoonotic mosaic.
538 *Clin. Microbiol. Rev.* **25**, 507–544 (2012).
539 7. Roncalli, R. Tracing the history of heartworms: A 400 year perspective. in *Recent*
540 *Advances in Heartworm Disease: AHS Symposium* (1998). doi:10.5555/20053201369.
541 8. Moroni, B. *et al.* Dirofilaria immitis in wolves recolonizing northern Italy: are wolves
542 competent hosts? *Parasit. Vectors* **13**, 482 (2020).
543 9. Aher, A. M. *et al.* Prevalence, genetic analyses, and risk factors associated with
544 heartworm (*Dirofilaria immitis*) in wild coyotes (*Canis latrans*) from Florida, USA. *J.*
545 *Wildl. Dis.* **52**, 785–792 (2016).
546 10. Wang, X. & Tedford, R. H. *Dogs: Their Fossil Relatives and Evolutionary History*.

(Columbia University Press, New York, NY, 2010).

11. Villmoare, B. *et al.* Paleoanthropology. Early Homo at 2.8 Ma from Ledi-Geraru, afar, Ethiopia. *Science* **347**, 1352–1355 (2015).

12. Bergström, A. *et al.* Grey wolf genomic history reveals a dual ancestry of dogs. *Nature* **607**, 313–320 (2022).

13. Freedman, A. H. *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10**, e1004016 (2014).

14. Botigué, L. R. *et al.* Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* **8**, 16082 (2017).

15. Soghigian, J. *et al.* Phylogenomics reveals the history of host use in mosquitoes. *Nat. Commun.* **14**, 6252 (2023).

16. Sallé, G. *et al.* The global diversity of *Haemonchus contortus* is shaped by human intervention and climate. *Nat. Commun.* **10**, 4811 (2019).

17. Doyle, S. R. *et al.* Population genomics of ancient and modern *Trichuris trichiura*. *Nat. Commun.* **13**, 3888 (2022).

18. Small, S. T. *et al.* Human Migration and the Spread of the Nematode Parasite *Wuchereria bancrofti*. *Mol. Biol. Evol.* **36**, 1931–1941 (2019).

19. Power, R. I., Doyle, S. R. & Šlapeta, J. Whole genome amplification and sequencing of individual *Dirofilaria immitis* microfilariae. *Exp. Parasitol.* **263-264**, 108806 (2024).

20. Choi, Y.-J. *et al.* Genomic diversity in *Onchocerca volvulus* and its *Wolbachia* endosymbiont. *Nat Microbiol* **2**, 16207 (2016).

21. Gandasegui, J. *et al.* Genome structure and population genomics of the canine heartworm *Dirofilaria immitis*. *Int. J. Parasitol.* (2023) doi:10.1016/j.ijpara.2023.07.006.

22. Penezić, A., Selaković, S., Pavlović, I. & Ćirović, D. First findings and prevalence of

572 adult heartworms (*Dirofilaria immitis*) in wild carnivores from Serbia. *Parasitol. Res.*
573 **113**, 3281–3285 (2014).
574 23. Ionică, A. M., Deak, G., Boncea, R., Gherman, C. M. & Mihalca, A. D. The European
575 badger as a new host for *Dirofilaria immitis* and an update on the distribution of the
576 heartworm in wild carnivores from Romania. *Pathogens* **11**, 420 (2022).
577 24. McCall, J. W., Genchi, C., Kramer, L. H., Guerrero, J. & Venco, L. Heartworm disease
578 in animals and humans. *Adv. Parasitol.* **66**, 193–285 (2008).
579 25. Molnár, V. *et al.* Autochthonous *Dirofilaria immitis* infection in a ferret with aberrant
580 larval migration in Europe. *J. Small Anim. Pract.* **51**, 393–396 (2010).
581 26. vonHoldt, B. M. *et al.* Whole-genome sequence analysis shows that two endemic
582 species of North American wolf are admixtures of the coyote and gray wolf. *Sci. Adv.* **2**,
583 e1501714 (2016).
584 27. Small, S. T., Tisch, D. J. & Zimmerman, P. A. Molecular epidemiology, phylogeny and
585 evolution of the filarial nematode *Wuchereria bancrofti*. *Infect. Genet. Evol.* **28**, 33–43
586 (2014).
587 28. Chavez, D. E. *et al.* Comparative genomics uncovers the evolutionary history,
588 demography, and molecular adaptations of South American canids. *Proc. Natl. Acad.*
589 *Sci. U. S. A.* **119**, e2205986119 (2022).
590 29. Leonard, J. A. *et al.* Megafaunal extinctions and the disappearance of a specialized wolf
591 ecomorph. *Curr. Biol.* **17**, 1146–1150 (2007).
592 30. Hofreiter, M. Pleistocene extinctions: haunting the survivors. *Curr. Biol.* **17**, R609–11
593 (2007).
594 31. Jiangzuo, Q. *et al.* Middle Pleistocene *Xenocyon lycaonoides* Kretzoi, 1938 in
595 northeastern China and the evolution of Xenocyon-Lycaon lineage. *Hist. Biol.* 1–13
596 (2022).

32. Rook, L. The wide ranging genus Eucyon Tedford & Qiu, 1996 (Mammalia, Carnivora, Canidae, Canini) in the Mio-Pliocene of the Old World. **31**, 723–741 (2009).

33. Kukla, G. J. *et al.* Last interglacial climates. *Quat. Res.* **58**, 2–13 (2002).

34. Batchelor, C. L. *et al.* The configuration of Northern Hemisphere ice sheets through the Quaternary. *Nat. Commun.* **10**, 3713 (2019).

35. Koblmüller, S. *et al.* Whole mitochondrial genomes illuminate ancient intercontinental dispersals of grey wolves (*Canis lupus*). *J. Biogeogr.* **43**, 1728–1738 (2016).

36. Koungoulos, L. G., Hulme-Beaman, A., Fillios, M. & Willandra Lakes Region World Heritage Aboriginal Advisory Group. Phenotypic diversity in early Australian dingoes revealed by traditional and 3D geometric morphometric analysis. *Sci. Rep.* **14**, 21228 (2024).

37. Fillios, M. A. & Taçon, P. S. C. Who let the dogs in? A review of the recent genetic evidence for the introduction of the dingo to Australia and implications for the movement of people. *J. Archaeol. Sci. Rep.* **7**, 782–792 (2016).

38. McKeever, B., Podadera, J. M., Beijerink, N. J. & Šlapeta, J. Suspect 'baggage canine heartworm' case: canine heartworm disease in a dog from Sydney, New South Wales. *Aust. Vet. J.* **99**, 359–362 (2021).

39. Ensminger, J. J. From hunters to hell hounds: the dogs of Columbus and transformations of the human-canine relationship in the early Spanish Caribbean. *Colon. Latin Am. Rev.* **31**, 354–380 (2022).

40. Symcox, G. & Sullivan, B. Introduction: Columbus—the man, the voyages, the legacy. in *Christopher Columbus and the Enterprise of the Indies* 1–40 (Palgrave Macmillan US, New York, 2005).

41. Welker, M. H. & Dunham, R. Exploring the introduction of European dogs to North America through shoulder height. *Intl J of Osteoarchaeology* **29**, 325–334 (2019).

622 42. Lau, D. C.-W. *et al.* Whole-genome reference of from Australia to determine single
623 nucleotide polymorphisms associated with macrocyclic lactone resistance in the USA.
624 *Curr Res Parasitol Vector Borne Dis* **1**, 100007 (2021).
625 43. Godel, C. *et al.* The genome of the heartworm, *Dirofilaria immitis*, reveals drug and
626 vaccine targets. *FASEB J.* **26**, 4650–4661 (2012).
627 44. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis
628 results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048
629 (2016).
630 45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
631 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
632 46. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat.*
633 *Biotechnol.* **35**, 316–319 (2017).
634 47. McIntyre, J. *et al.* Chromosomal genome assembly resolves drug resistance loci in the
635 parasitic nematode *Teladorsagia circumcincta*. *bioRxiv* 2024.07.22.603261 (2024)
636 doi:10.1101/2024.07.22.603261.
637 48. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**,
638 3094–3100 (2018).
639 49. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker,*
640 *GATK, and WDL in Terra*. (O'Reilly Media, 2020).
641 50. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008
642 (2021).
643 51. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast
644 processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
645 52. Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P. & Marth, G. T.
646 BamTools: a C++ API and toolkit for analyzing and managing BAM files.

647 *Bioinformatics* **27**, 1691–1692 (2011).
648 53. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
649 genomic features. *Bioinformatics* **26**, 841–842 (2010).
650 54. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158
651 (2011).
652 55. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal
653 component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
654 56. Korunes, K. L. & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and
655 divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
656 57. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture
657 proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
658 58. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak:
659 a program for identifying clustering modes and packaging population structure
660 inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
661 59. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from
662 genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
663 60. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093
664 (2012).
665 61. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and
666 richer datasets. *Gigascience* **4**, 7 (2015).
667 62. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population
668 history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).

669

670

## Acknowledgements

The authors gratefully acknowledge Julie Austin, Iris Cheadle, Constantin Constantinoiu, Carol Esson, Eloise Fenton, Linda Hayes, Tessa Mackie, Yissu Martinez, Sandi McClintock, Ilze Nel, Tony Phillis, Duncan Smith, and Jim Taylor for their generous provision of heartworm specimens for this study. Sampling from the University of Queensland was supported by the University of Queensland, School of Veterinary Science Bequest and Donation Funds for Research (Stone Memorial Fund). We thank Javier Gandasegui for their foundational bioinformatics work that facilitated this work. We acknowledge the Pathogen Informatics group at the Wellcome Sanger Institute and the Sydney Informatics Hub for bioinformatics support. RIP is supported by an Australian Government Research Training Program Stipend and The Jean Walker Trust Fellowship. RIP's collaborative research exchange was supported by the University of Sydney's Eric Horatio Scholarship, the University of Sydney's Postgraduate Research Support Scheme, and the Australian Society for Parasitology's Researcher Exchange, Training and Travel Award. SRD is supported by a UKRI Future Leaders Fellowship [MR/T020733/1] and the Wellcome Trust (UK) through core funding to the Wellcome Sanger Institute (UK) [206194]. This work was funded by the Canine Research Foundation, Dogs Victoria, Australia and the Australian Companion Animal Health Foundation Research Grant.

# Figures

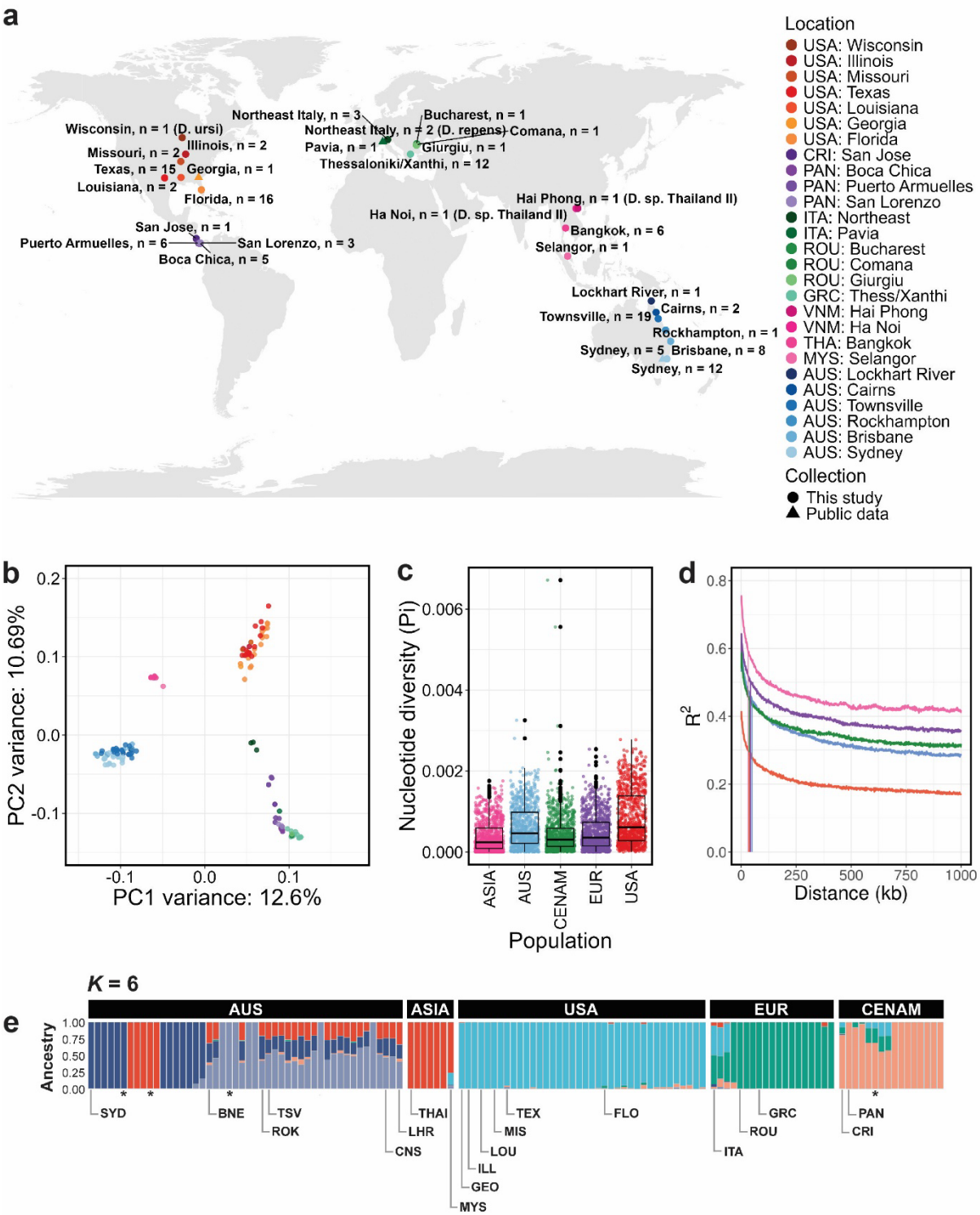



**Fig. 1: Global sampling and population structure of heartworms. a,** World map showing the location of adult heartworm specimens used in the study. Adult *Dirofilaria ursi*, *D. repens*, and *D.* sp. 'Thailand II' were included as outgroup samples. Abbreviations: USA = United

States of America (Georgia: GEO, Illinois: ILL, Louisiana: LOU, Missouri: MIS, Texas: TEX, Florida: FLO); CRI = Costa Rica; PAN = Panama; ITA = Italy; ROU = Romania; GRC = Greece; VNM = Vietnam; THA = Thailand; MYS = Malaysia; AUS = Australia (Sydney: SYD, Brisbane: BNE, Rockhampton: ROK, Townsville: TSV, Cairns: CNS, Lockhart River: LHR). **b,** Principal component analysis (PCA) of 196,550 nuclear single nucleotide polymorphisms (SNPs) from 124 heartworms. **c,** Box plots showing the nucleotide diversity (Pi) distribution within broad geographical regions, with each point representing a 100 kb sliding window. **d,** Linkage disequilibrium (LD) decay plot for the heartworm cohort, grouped by broad geographic region. Vertical lines show the distance between SNPs where the LD is half the maximum value. **e,** Admixture inferred using NGSadmix with six clusters (K). Each vertical bar represents an individual sample, with colours representing the clusters. Labels at the top of the bar charts indicate the broad geographic origins of samples, while those at the bottom show their more specific locations. An asterisk (*) denotes a replicate of the preceding sample.

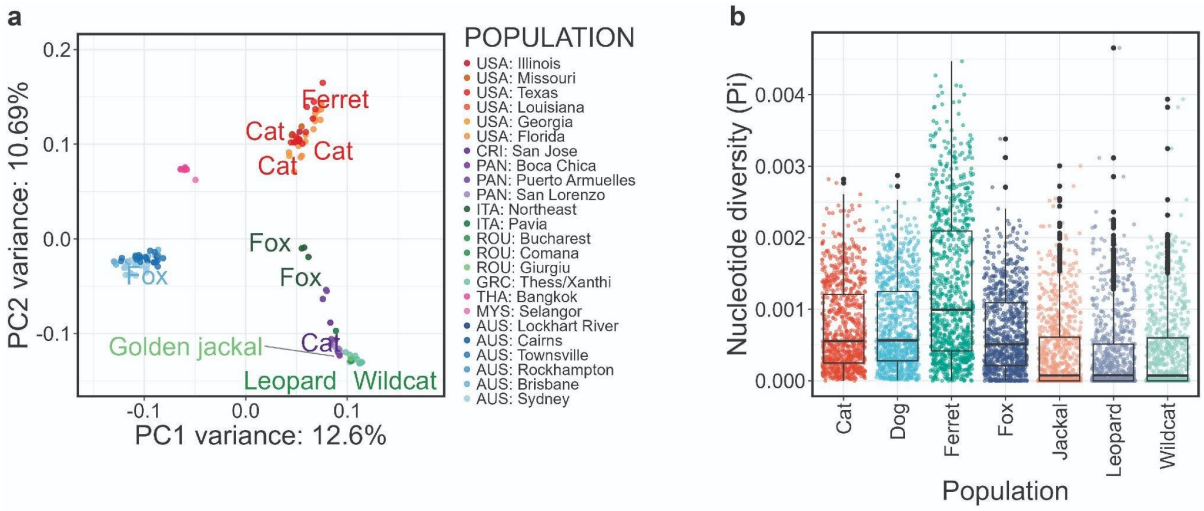



**Fig. 2: Heartworm variation between carnivoran hosts. a,** Principal component analysis (PCA) of heartworm nuclear diversity from Fig. 1b, with non-dog hosts labelled. The geographic origin of samples is indicated on the right (i.e. COUNTRY: city or region). Country abbreviations: USA = United States of America; CRI = Costa Rica; PAN = Panama; ITA = Italy; ROU = Romania; GRC = Greece; VNM = Vietnam; THA = Thailand; MYS = Malaysia; AUS = Australia. **b,** Box plots show the distribution of nucleotide diversity (Pi) within host populations. Each data point represents a sliding window size of 100 kb.

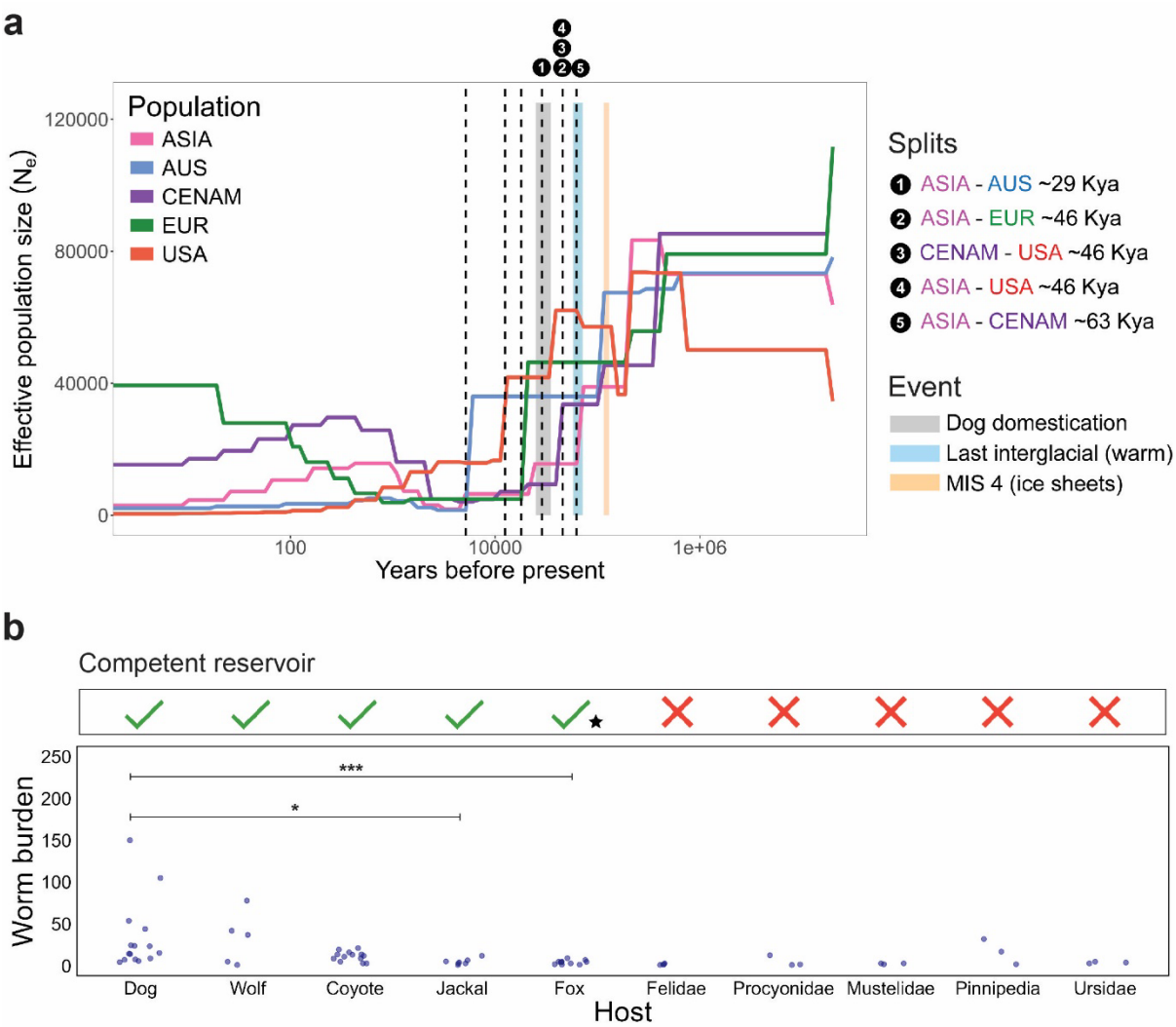



**Fig. 3: Demographic history and host competency for heartworms. a,** Effective population size history of heartworm populations inferred using SMC++ based on a four-year generation time. Each line represents the effective population size of heartworms on each continent over time. The grey box spans ~14–40 kya, highlighting the previously suggested period of dog domestication. Vertical dotted lines and associated numbers represent divergence times between populations. Details of these population splits and their timing are provided on the right. Abbreviations: AUS = Australia; CENAM = Central America; EUR = Europe; USA = United States of America. **b,** Support for the heartworm life cycle and adult worm burdens in carnivoran hosts. The top panel highlights hosts that are frequently microfilaremic and are hence considered competent reservoirs for heartworms. A black star indicates cases where

728 microfilaremia has been reported in the host, but they are generally considered to pose a low
729 risk of being competent reservoirs. The bottom panel is a plot showing adult heartworm
730 burdens across various hosts. Each point represents data from a study, sourced from the
731 literature (Supplementary Table 2). Significant differences in worm burdens among competent
732 reservoirs were assessed using a Kruskal-Wallis test with Dunn's multiple comparisons. P-
733 values ≤ 0.05 were considered statistically significant (*: ≤ 0.05, **: ≤ 0.01, ***: ≤ 0.001).

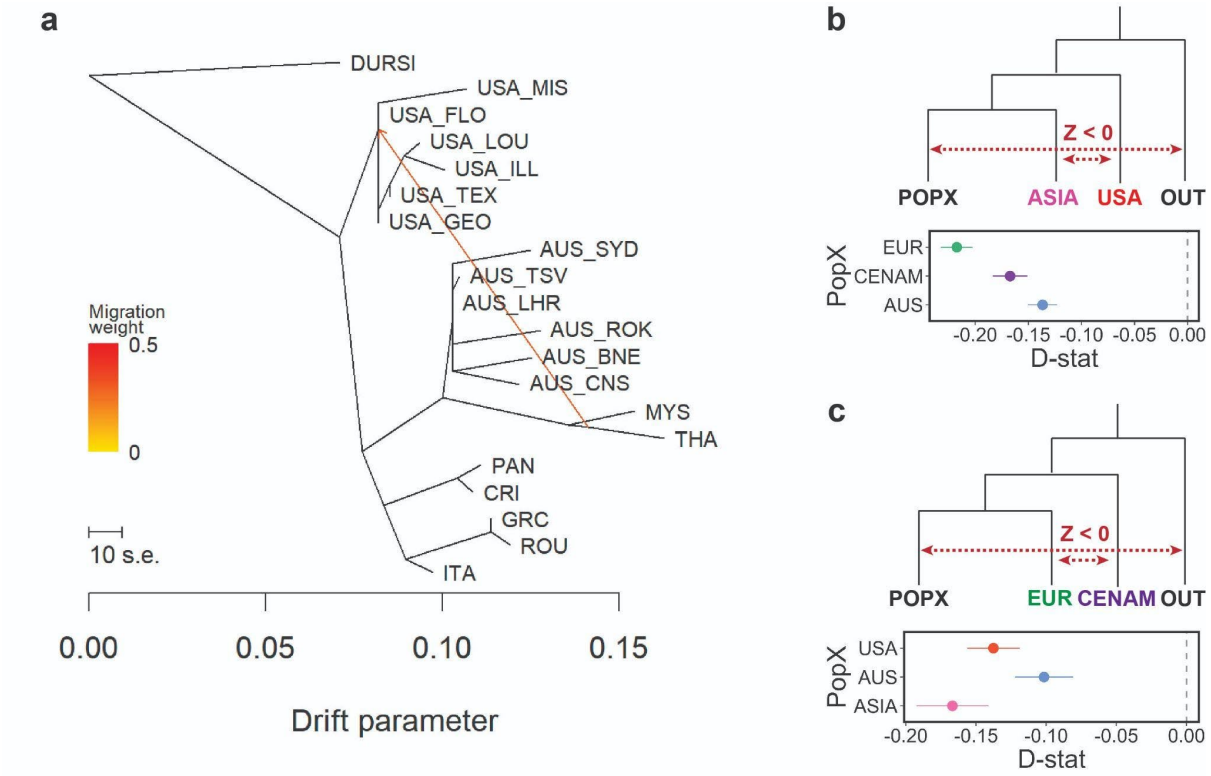
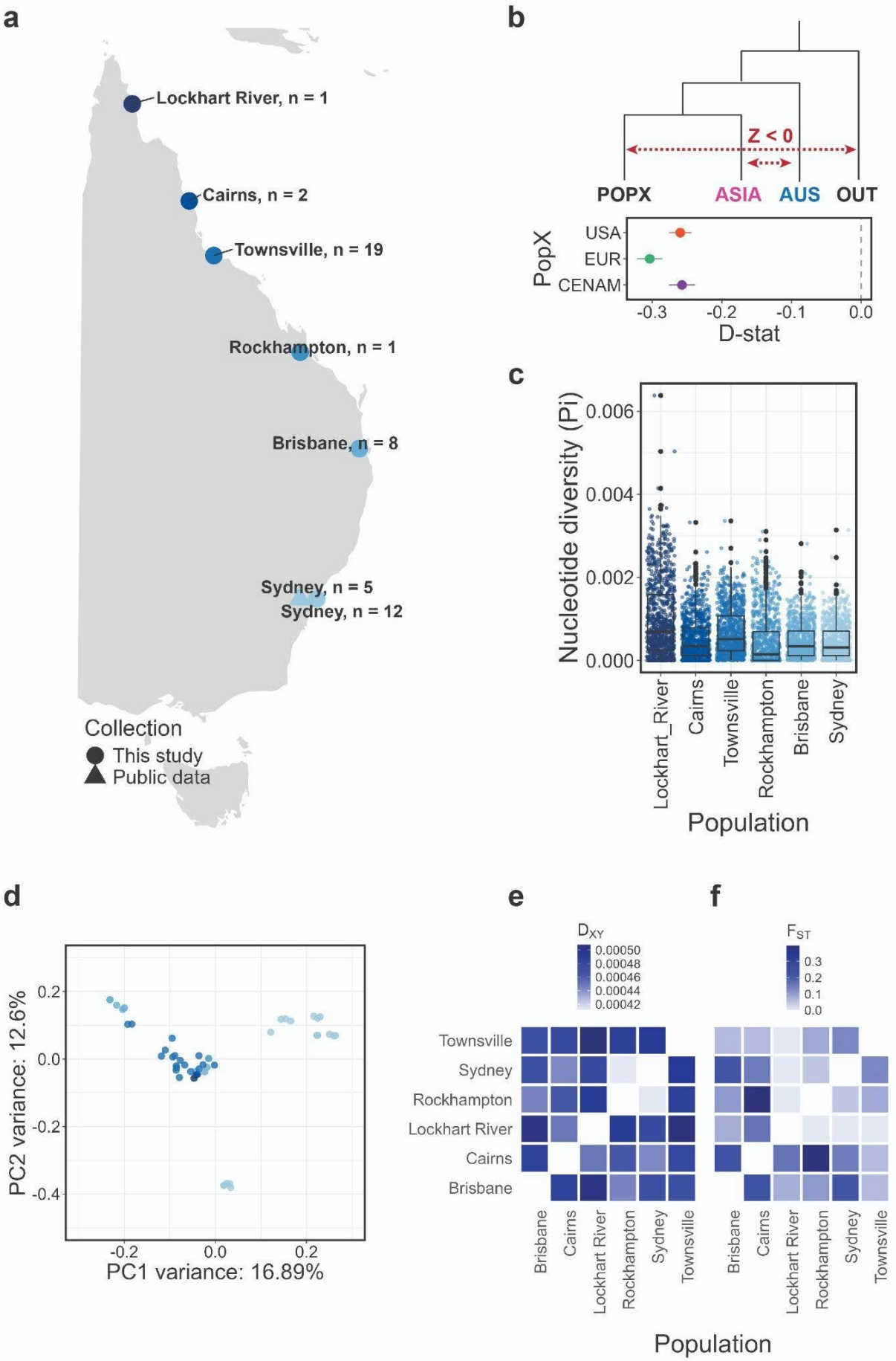



**Fig. 4: Transcontinental admixture in heartworms. a,** Maximum-likelihood tree of heartworm generated using TreeMix, with *Dirofilaria ursi* (DURSI) as an outgroup and one migration edge. **b,c,** *D*-statistics for hypothetical allele-sharing scenarios between heartworm populations using Admixtools qpDstat with 21,088 single nucleotide polymorphisms (SNPs) and *D. ursi* as the outgroup (OUT). *D*-statistics test for admixture between four populations (W, X, Y, Z). A Z-score (*D*-statistic / standard error) is calculated, which provides information about the direction of gene flow (red dotted lines) (Z-score > 0 = gene flow between either [W and Y], or [X and Z]; Z-score < 0 = gene flow between either [W and Z] or [X and Y]). For all data presented, the Z score was significant (|Z| > 3). Error bars show the standard error.


a
Lockhart River, n = 1
Cairns, n = 2
Townsville, n = 19
Rockhampton, n = 1
Brisbane, n = 8
Sydney, n = 5
Sydney, n = 12
Collection
This study
Public data
b
Z < 0
POPX
ASIA
AUS
OUT
PopX
USA
EUR
CENAM
-0.3
-0.2
-0.1
0.0
D-stat
c
Nucleotide diversity (Pi)
0.006
0.004
0.002
0.000
Lockhart_River
Cairns
Townsville
Rockhampton
Brisbane
Sydney
Population
d
PC2 variance: 12.6%
0.2
0.0
-0.2
-0.4
-0.2
0.0
0.2
PC1 variance: 16.89%
e
$D_{XY}$
0.00050
0.00048
0.00046
0.00044
0.00042
f
$F_{ST}$
0.3
0.2
0.1
0.0
Townsville
Sydney
Rockhampton
Lockhart River
Cairns
Brisbane
Brisbane
Cairns
Lockhart River
Rockhampton
Sydney
Townsville
Population

744

745 **Fig. 5: Gene flow and population structure of Australian heartworms. a,** Distribution of
746 adult heartworm samples ($n$ = 48) collected along the east coast of Australia. **b,** $D$-statistics
747 showing excessive allele sharing between Australia (AUS) and ASIA relative to that between
748 AUS and other regions (PopX). Admixtools qpDstat was used to analyse 21,088 single
749 nucleotide polymorphisms (SNPs), with *Dirofilaria ursi* as the outgroup (OUT). $D$-statistics
750 test for admixture between four populations (W, X, Y, Z). A Z-score ($D$-statistic / standard
751 error) is calculated, which provides information about the direction of gene flow (red dotted
752 lines) (Z-score > 0 = gene flow between [W and Y], or [X and Z]; Z-score < 0 = gene flow
753 between [W and Z] or [X and Y]). For all data presented, the Z score was significant ($|Z| > 3$).
754 Error bars show the standard error. **c,** Box plots showing the nucleotide diversity (Pi)
755 distribution from dog hosts per city. Each data point represents a 100 kb sliding window. **d,**
756 Principal component analysis (PCA) of 116,492 nuclear SNPs in the Australian cohort. **e**,
757 Absolute nucleotide divergence ($D_{XY}$) and **f,** genetic differentiation ($F_{ST}$) between Australian
758 cities using a 100 kb sliding window. Only heartworms from dog hosts were included in the
759 analysis.

760

761

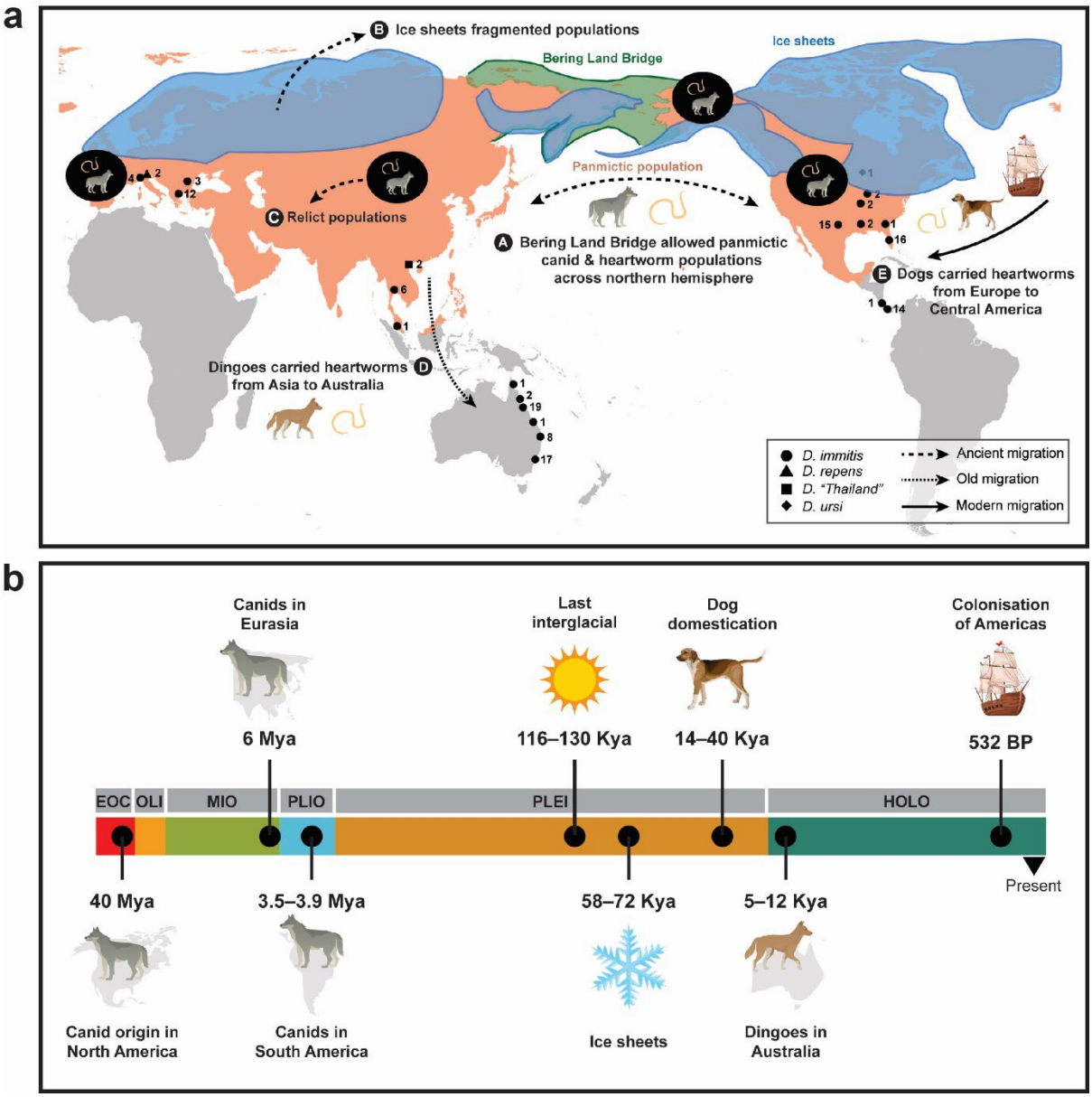
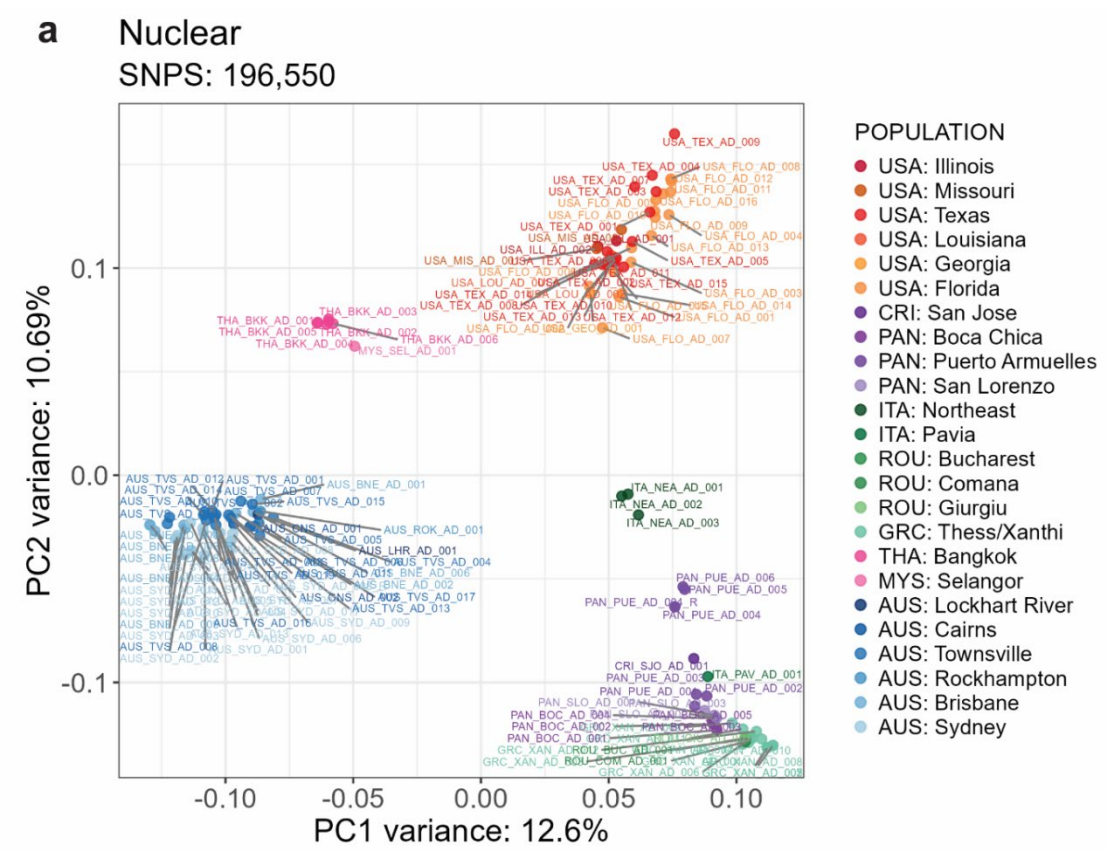


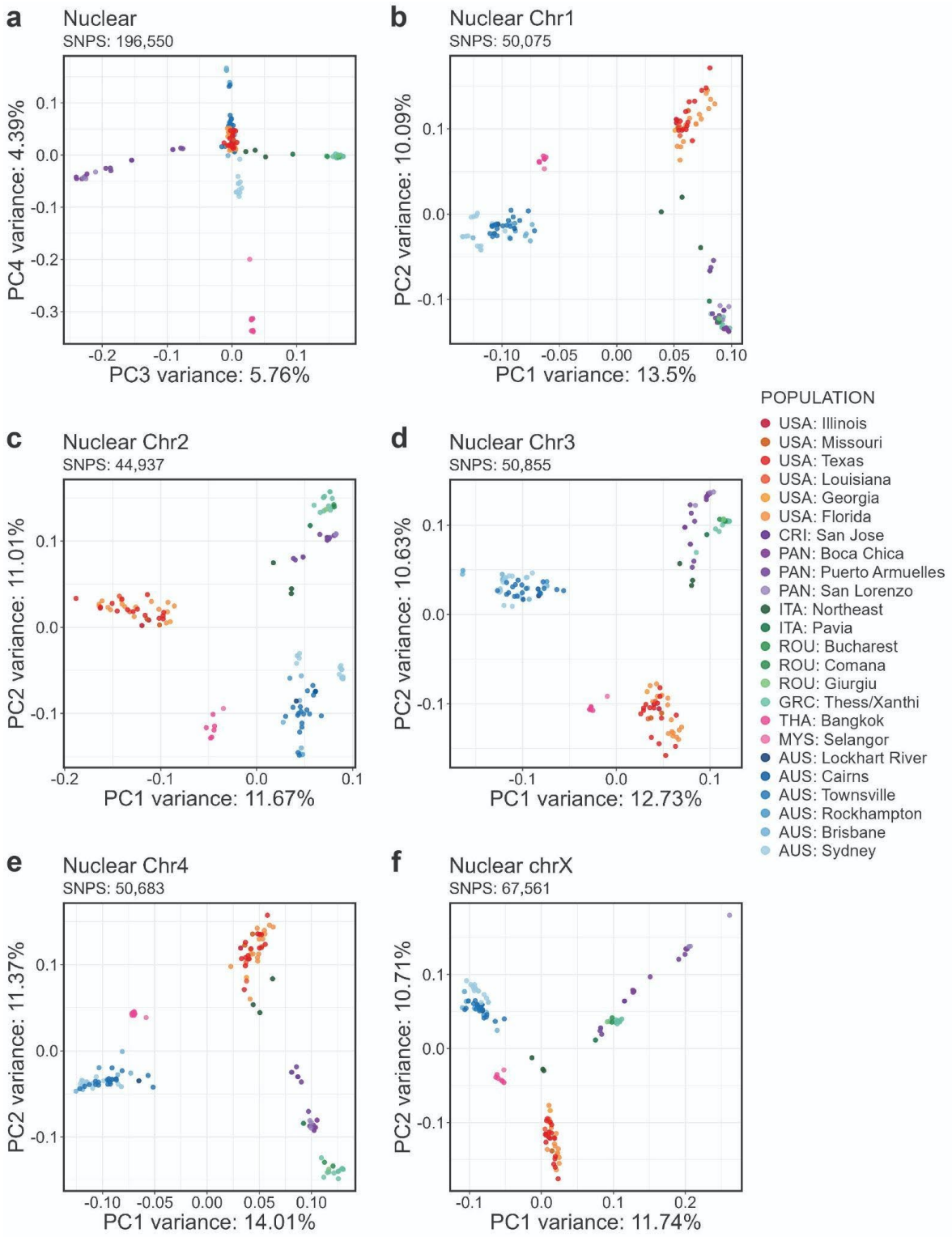

**Fig. 6: A new evolutionary paradigm of heartworms throughout history. a,** Hypothetical heartworm dispersal scenarios supported in the current study. Pink regions represent a panmictic population of ancient canids harbouring heartworms across North America and Eurasia. The Bering Land Bridge is shaded in green, and continental ice sheets are shown in blue. Events are in chronological order (A–E), with arrows illustrating ancient (dashed), old (dotted), and modern (solid) migration events. Small black shapes with adjacent numbers indicate the geographic origin of samples collected in this study and their sample size. The legend specifies the parasite species, including the heartworm *Dirofilaria immitis* ($n$ = 127)

and outgroups *D. repens* (*n* = 2), D. sp. "Thailand II" (*n* = 2), and *D. ursi* (*n* = 1). **b,** Timeline of key canid evolution and climatic events. Horizontal bars represent Epochs: HOLO = Holocene; PLEI = Pleistocene; PLIO = Pliocene; MIO = Miocene; OLI = Oligocene; EOC = Eocene. The time scale was log-transformed to enhance the visualisation of events.

## Extended Data Figures

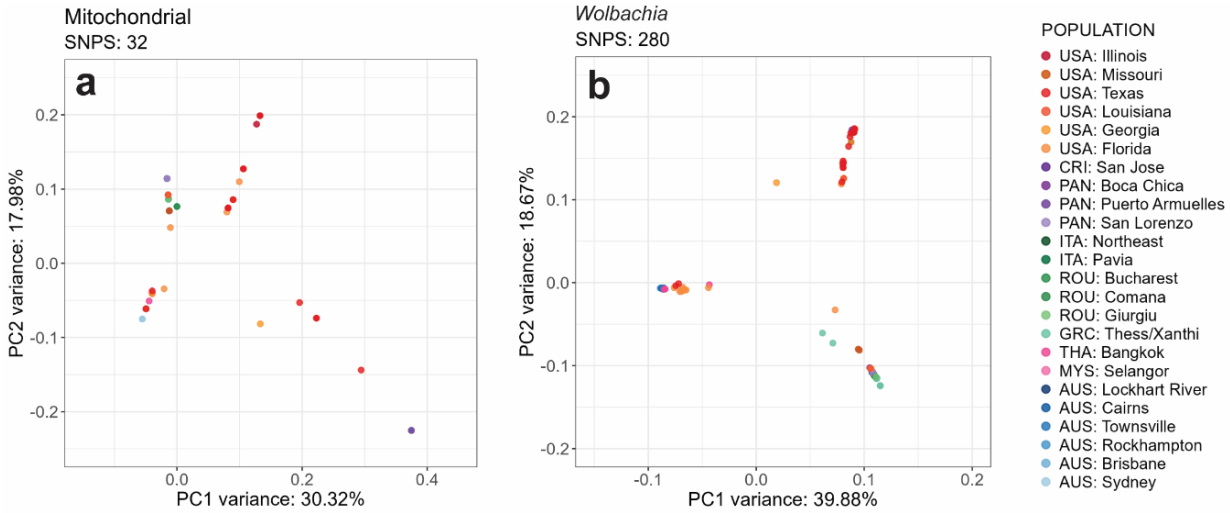



**Extended Data Fig. 1: Principal component analysis (PCA) of 196,550 nuclear single nucleotide polymorphisms (SNPs) from 124 heartworms with labelled metadata.** Samples were collected from Australia (AUS), the USA, Panama (PAN), Greece (GRC), Thailand (THA), Italy (ITA), Romania (ROU), Costa Rica (CRI), and Malaysia (MYS). Replicate samples were included in the analyses. The legend on the right indicates the geographic origin of samples (i.e. COUNTRY: city or region). Principal component 1 (PC1) and principal component 2 (PC2) are shown on the x and y axes, explaining 12.6% and 10.69% of the variation, respectively. These figures are identical to Fig. 1b but include sample labels.

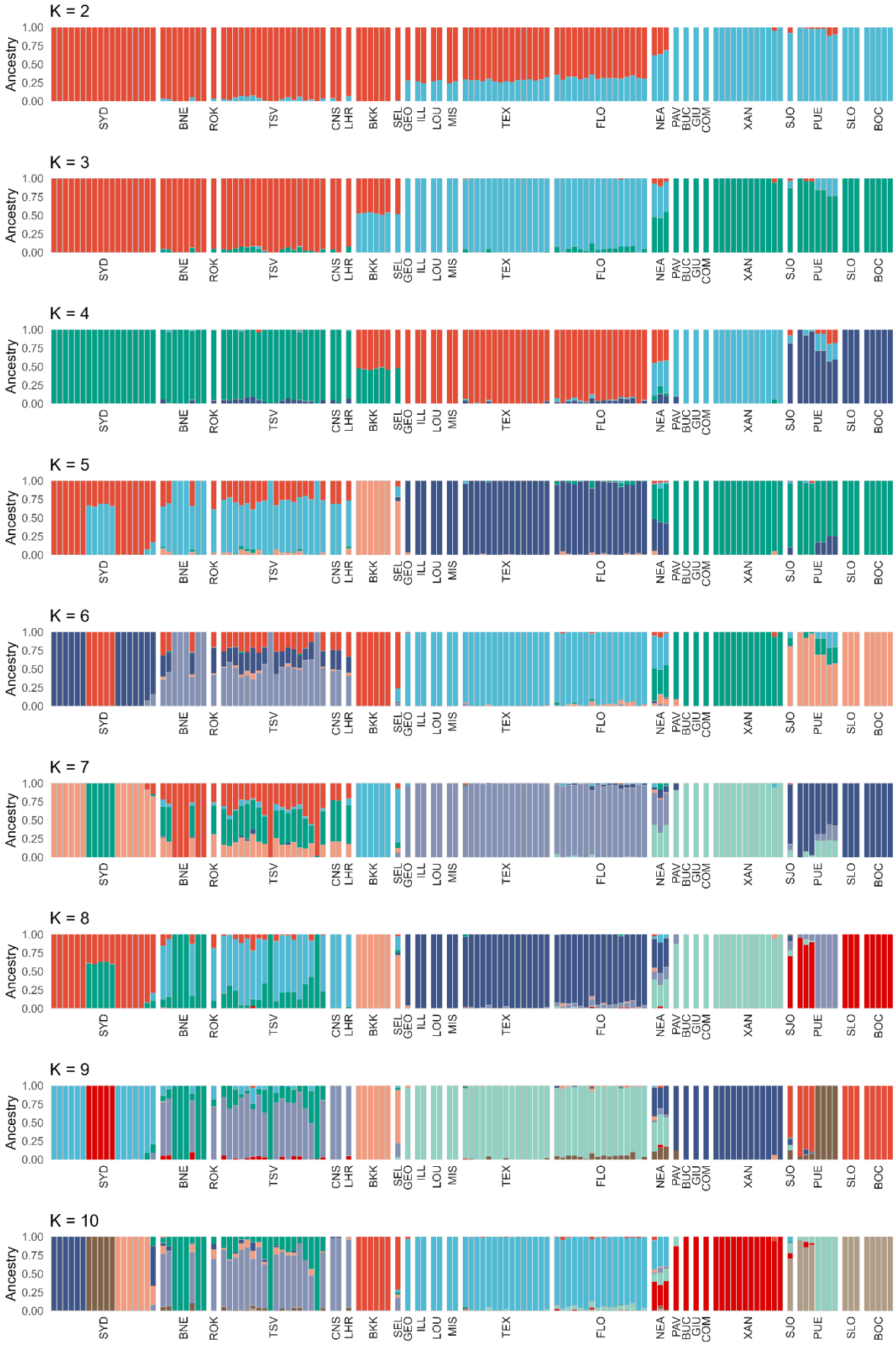



**Extended Data Fig. 2: Principal component analysis (PCA) of 124 heartworms from around the world based on the nuclear genome.** Samples were collected from Australia (AUS), the USA, Panama (PAN), Greece (GRC), Thailand (THA), Italy (ITA), Romania (ROU), Costa Rica (CRI), and Malaysia (MYS). Replicate samples were included in the

790 analyses. The legend on the right indicates the geographic origin of samples (i.e. COUNTRY:
791 city or region). The principal components (PCs) are shown on the x and y axes. **a,** PCA of
792 196,550 nuclear single nucleotide polymorphisms (SNPs) showing PC3 and PC4, which
793 explained 5.76% and 4.39% of the variation, respectively. **b–f,** PCAs of 44,937 – 67,561 SNPs
794 in chromosome 1, chromosome 2, chromosome 3, chromosome 4, and the sex-linked X-
795 chromosome of canine heartworm. PC1 and PC2 explained 11.67 – 14.01% and 10.09 –
796 11.37% of the variation, respectively.

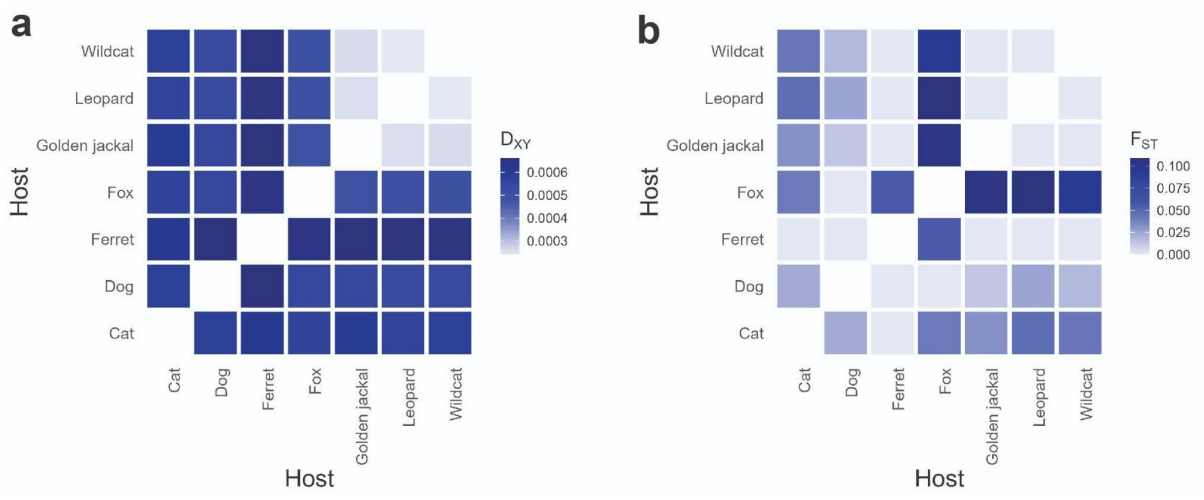



**Extended Data Fig. 3: Genetic relatedness of the mitochondria and *Wolbachia* endosymbionts of heartworms.** Abbreviations: USA = United States of America; CRI = Costa Rica; PAN = Panama; ITA = Italy; ROU = Romania; GRC = Greece; VNM = Vietnam; THA = Thailand; MYS = Malaysia; AUS = Australia. **a,** Principal component analysis (PCA) of 32 mitochondrial single nucleotide polymorphisms (SNPs) from 127 heartworms. Principal component 1 (PC1) and principal component 2 (PC2) explained 30.32% and 17.98% of the variation, respectively. **b,** PCA of 280 *Wolbachia* endosymbiont SNPs from 121 samples. PC1 and PC2 explained 39.88% and 18.67% of the variation, respectively. Replicate samples were included in all PCAs. The legend on the right indicates the geographic origin of samples (i.e. COUNTRY: city or region).

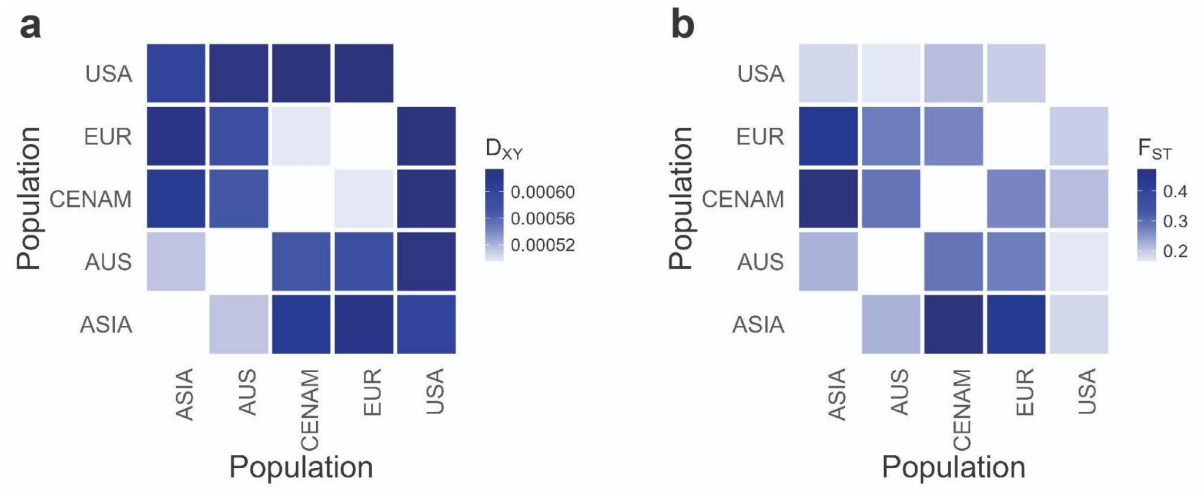

K = 2
K = 3
K = 4
K = 5
K = 6
K = 7
K = 8
K = 9
K = 10
Ancestry
1.00
0.75
0.50
0.25
0.00
SYD
BNE
ROK
TSV
CNS
LHR
BKK
SEL
GEO
ILL
LOU
MIS
TEX
FLO
NEA
PAV
BUC
GIU
COM
XAN
SJO
PUE
SLO
BOC

808

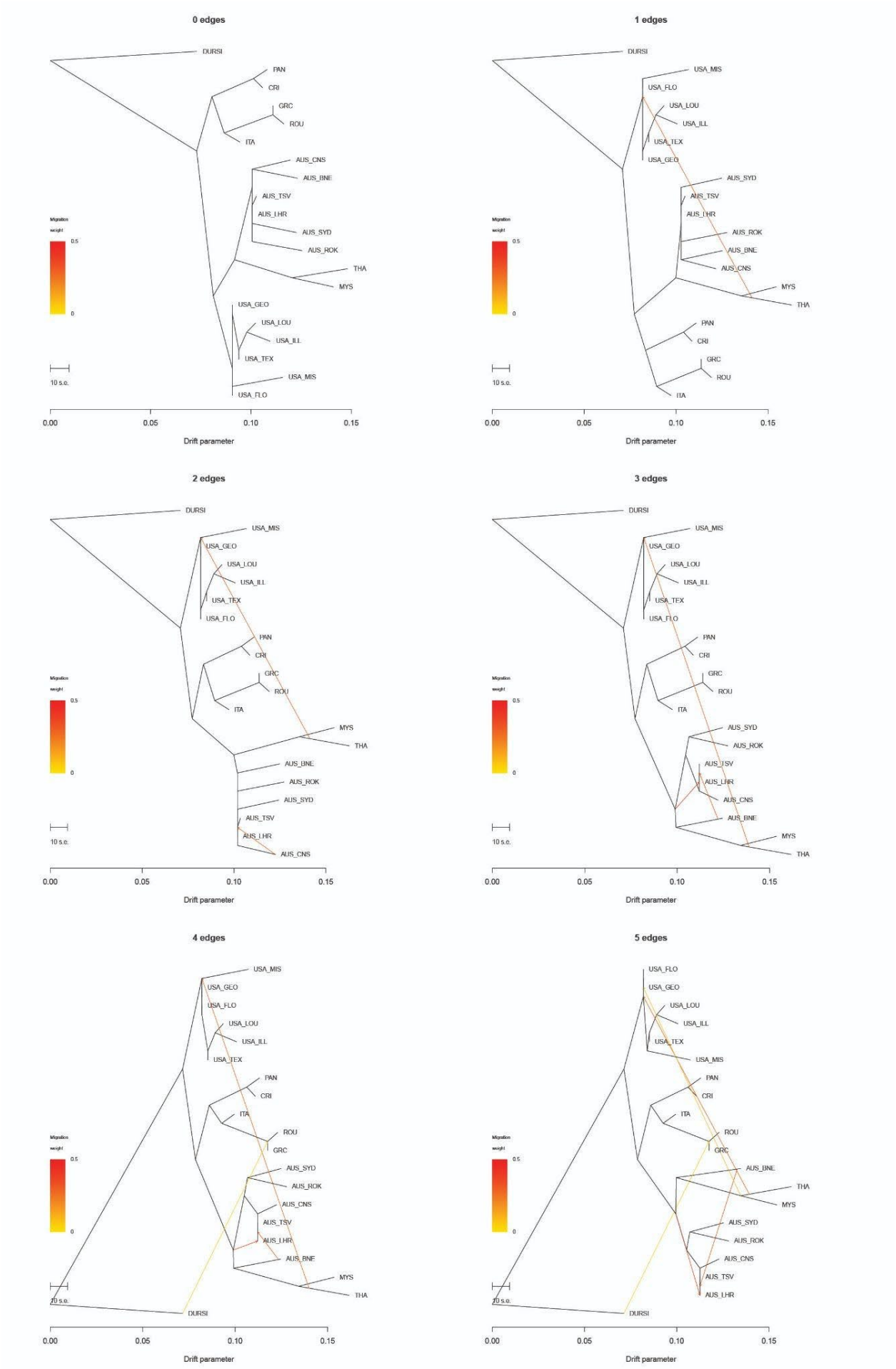
809 **Extended Data Fig. 4: Admixture analysis of adult heartworms.** Analysis was performed
810 in NGSadmix across a range of K values (2 – 10) (colours) for 124 heartworms (vertical bars).
811 Geographic origins are indicated at the bottom of each chart. Abbreviations: Australia (Sydney:
812 SYD, Brisbane: BNE, Rockhampton: ROK, Townsville: TSV, Cairns: CNS, Lockhart River:
813 LHR); Thailand (Bangkok: BKK); Malaysia (Selangor: SEL); USA (Georgia: GEO, Illinois:
814 ILL, Louisiana: LOU, Missouri: MIS, Texas: TEX, Florida: FLO); Italy (Northeast: NEA,
815 Pavia: PAV); Romania (Bucharest: BUC, Giurgiu: GIU, Comana: COM); Greece
816 (Thessaloniki/Xanthi: XAN); Costa Rica (San José: SJO); Panama (Puerto Armuelles: PUE,
817 San Lorenzo: SLO, Boca Chica: BOC).




**Extended Data Fig. 5: Population genetic summary statistics of heartworms collected from diverse host populations.** Heatmaps show **a,** Absolute nucleotide divergence ($D_{XY}$) and **b,** Genetic differentiation ($F_{ST}$) using a 100 kb sliding window.

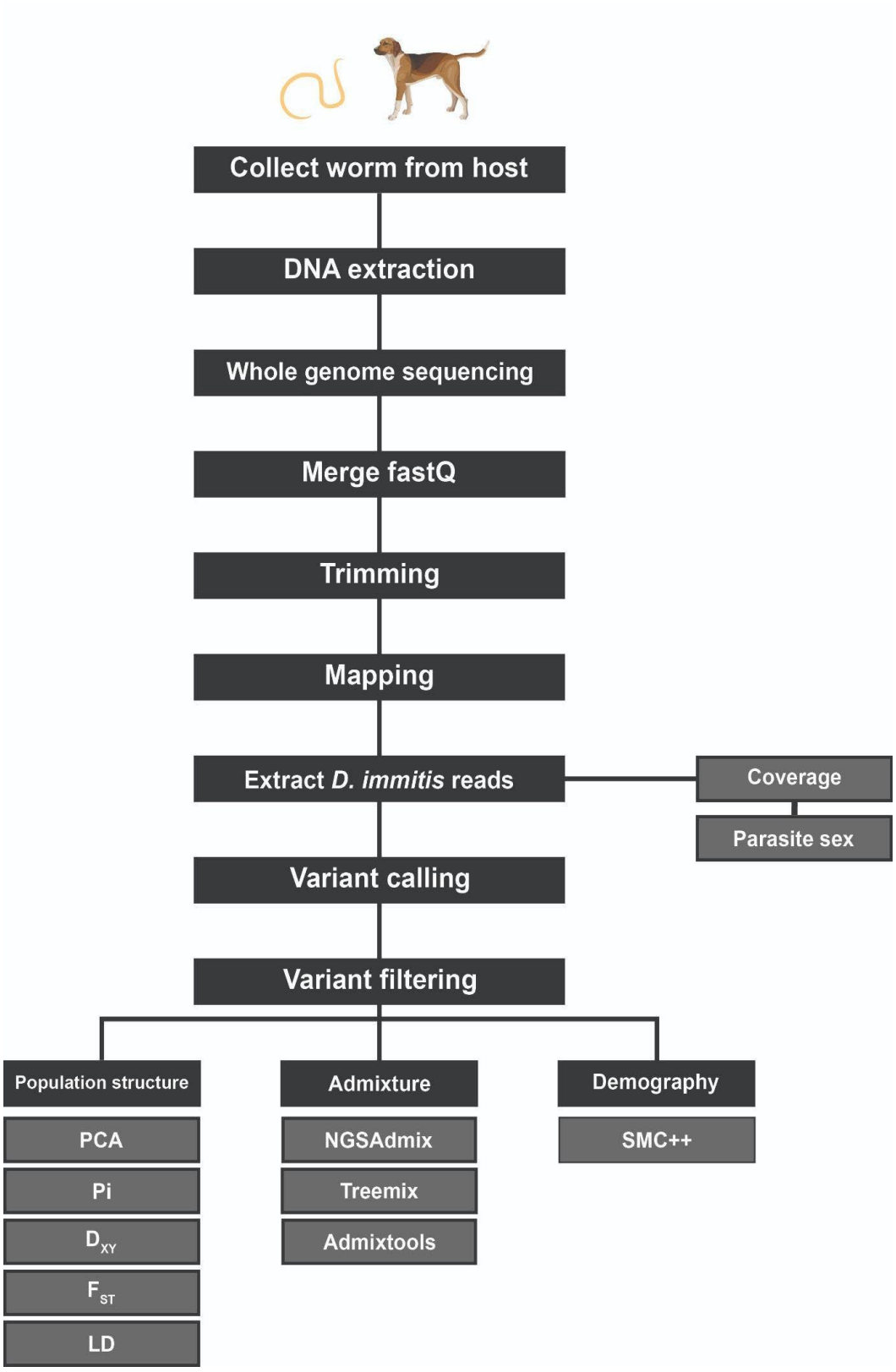



**Extended Data Fig. 6: Population genetic summary statistics of heartworms between broad geographical regions.** Heatmaps show **a,** Absolute nucleotide divergence ($D_{XY}$) and **b,** Genetic differentiation ($F_{ST}$) using a 100 kb sliding window. Region abbreviations: ASIA = Asia; AUS = Australia; CENAM = Central America; EUR = Europe; USA = United States of America.

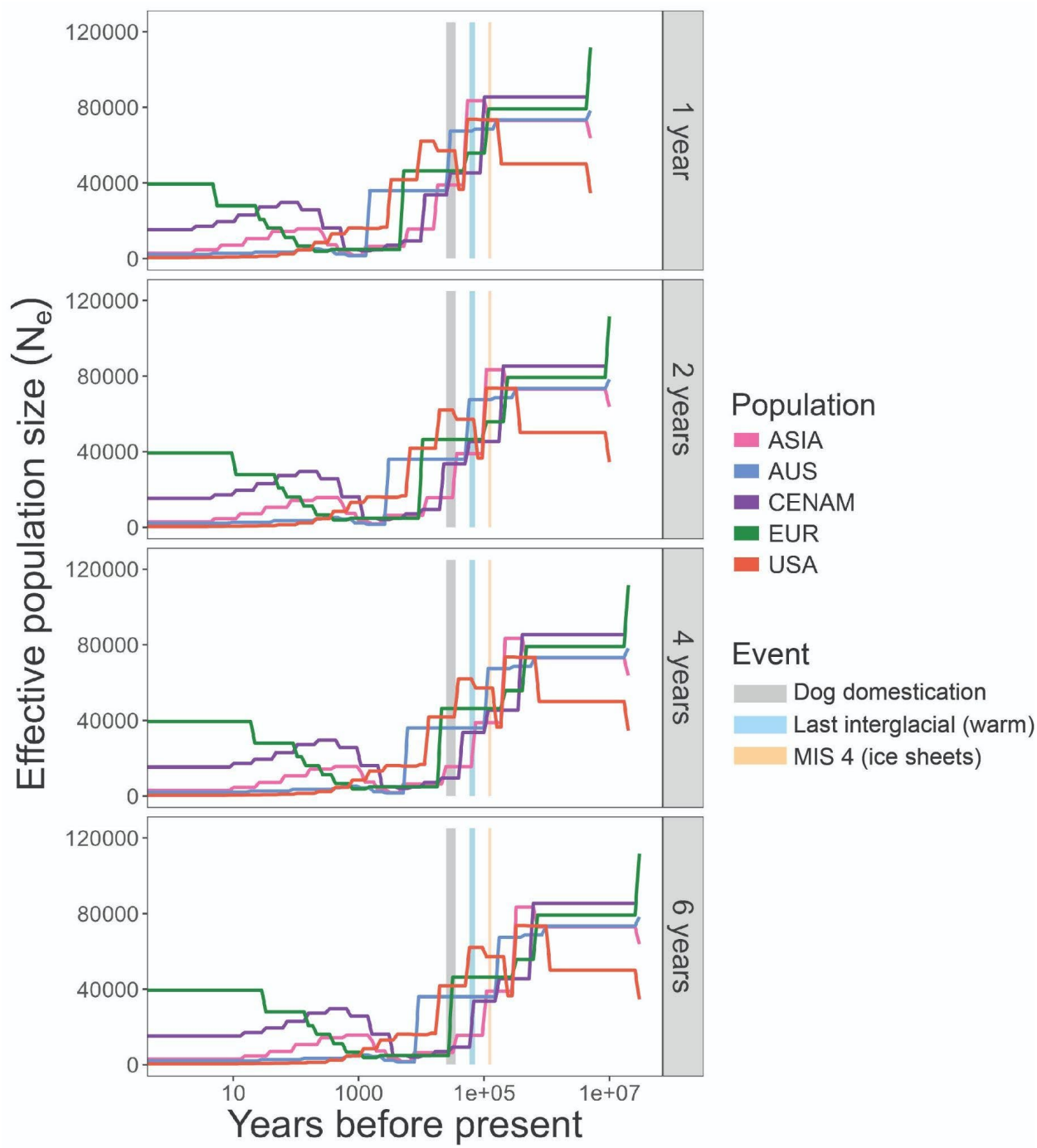

0 edges
1 edges
2 edges
3 edges
4 edges
5 edges
DURSI
PAN
CRI
GRC
ROU
ITA
AUS_CNS
AUS_BNE
AUS_TSV
AUS_LHR
AUS_SYD
AUS_ROK
THA
MYS
USA_GEO
USA_LOU
USA_ILL
USA_TEX
USA_MIS
USA_FLO
Migration weight
0.5
0
10 s.e.
0.00
0.05
0.10
0.15
Drift parameter

830

831 **Extended Data Fig. 7: TreeMix analysis of nuclear heartworm data.** Maximum likelihood
832 trees are shown with migration edges ranging from 0 to 5, using *Dirofilaria ursi* as an outgroup.
833 Coloured lines represent inferred migration events, with the colour indicating support levels.
834 Abbreviations: Australia: AUS (Sydney: SYD, Brisbane: BNE, Rockhampton: ROK,
835 Townsville: TSV, Cairns: CNS, Lockhart River: LHR); USA (Georgia: GEO, Illinois: ILL,
836 Louisiana: LOU, Missouri: MIS, Texas: TEX, Florida: FLO); Thailand: THA; Malaysia: MYS;
837 Italy: ITA; Romania: ROU; Greece: GRC; Costa Rica: CRI; Panama: PAN.



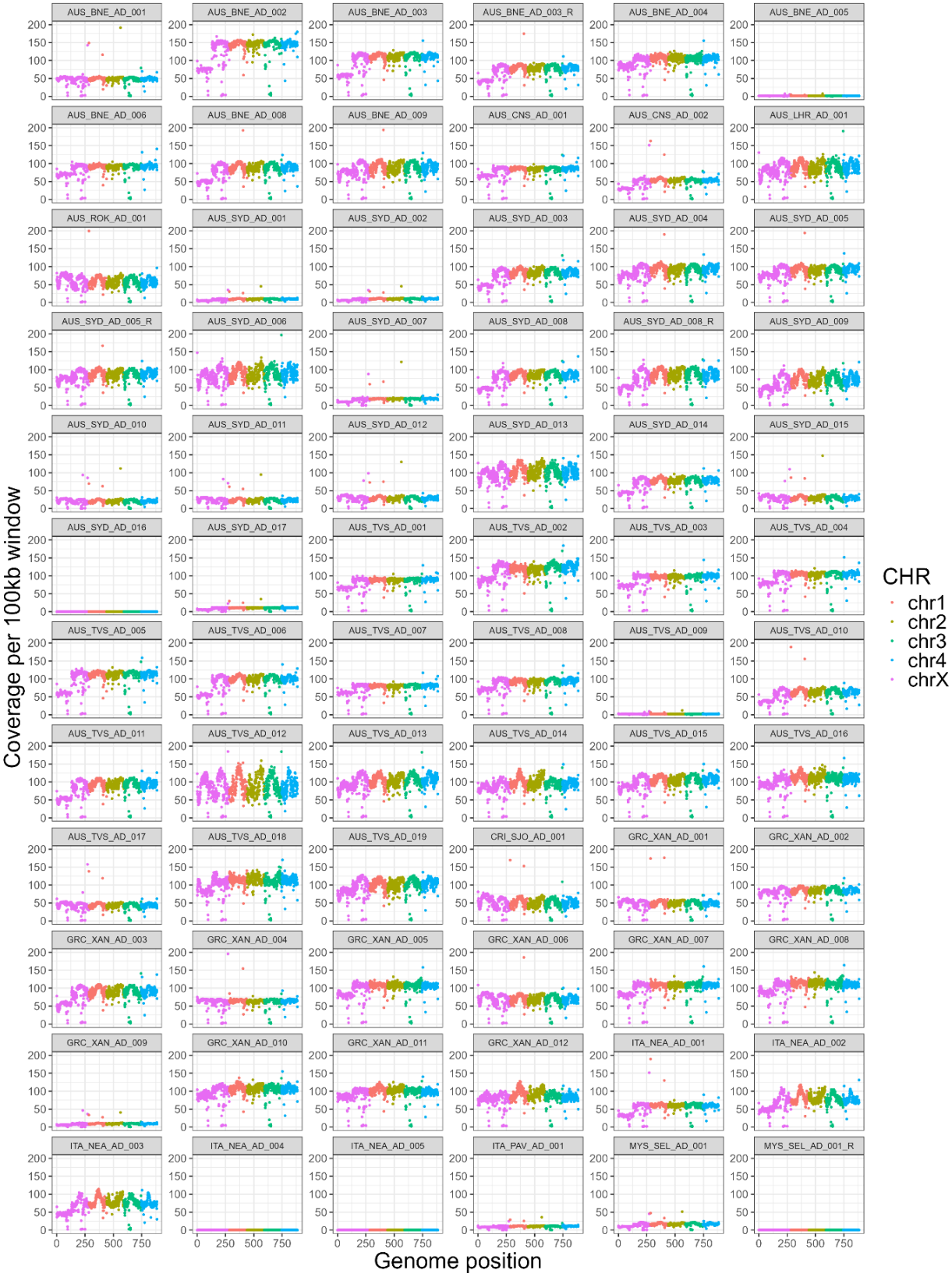

**Extended Data Fig. 8: Summary diagram of the population genomics workflow in this study.** Abbreviations of population structure analyses: PCA = Principal Component Analysis; Pi = nucleotide diversity; $D_{XY}$ = absolute nucleotide divergence; $F_{ST}$ = genetic differentiation.




**Extended Data Fig. 9: Demographic history analysis of heartworms from dogs using a range of generation times.** Size histories were inferred using SMC++, with each line representing a heartworm subpopulation. A range of generation times were tested for *Dirofilaria immitis* (i.e. 1, 2, 4, and 6 years). The grey box indicates previously suggested date ranges for dog domestication (~14–40 kya). Abbreviations: AUS = Australia; CENAM =

Central America; EUR = Europe; USA = United States of America; MIS 4 = Marine Isotype Stage 4.

# Supplementary Figures



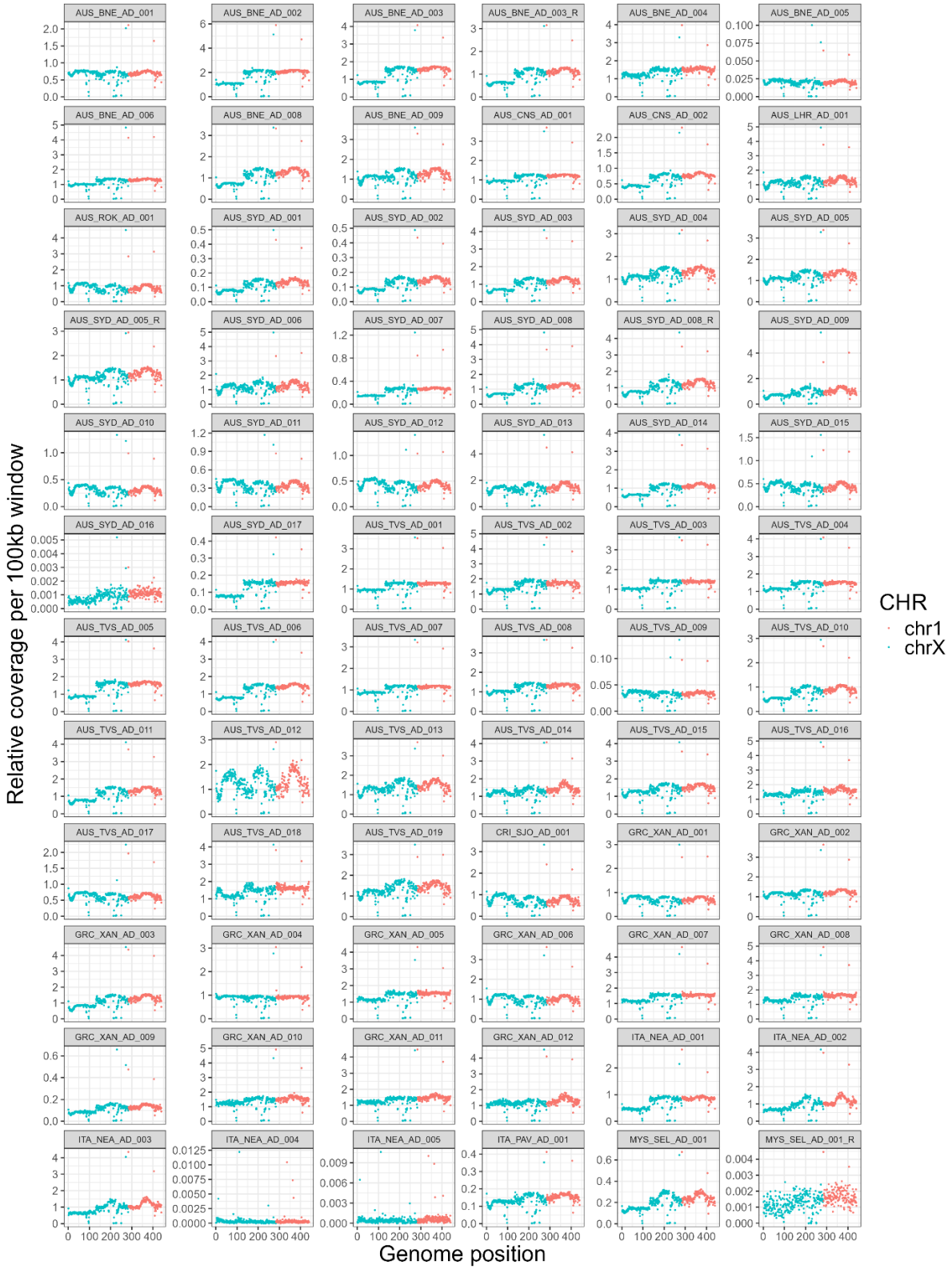
852 **Supplementary Fig. 1: Raw nuclear sequencing coverage per sample (Part 1/2).** Coverage
853 was calculated using 100 kb sliding windows for all mapped samples. Each chromosome is
854 shown in a different colour, with a legend on the right. The remaining samples are shown in
855 Supplementary Fig. 2.

PAN_BOC_AD_001
PAN_BOC_AD_002
PAN_BOC_AD_003
PAN_BOC_AD_004
PAN_BOC_AD_005
PAN_PUE_AD_001
PAN_PUE_AD_002
PAN_PUE_AD_003
PAN_PUE_AD_004
PAN_PUE_AD_004_R
PAN_PUE_AD_005
PAN_PUE_AD_006
PAN_SLO_AD_001
PAN_SLO_AD_002
PAN_SLO_AD_003
ROU_BUC_AD_001
ROU_COM_AD_001
ROU_GIU_AD_001
ROU_GIU_AD_001_R
THA_BKK_AD_001
THA_BKK_AD_002
THA_BKK_AD_003
THA_BKK_AD_004
THA_BKK_AD_005
THA_BKK_AD_006
USA_FLO_AD_001
USA_FLO_AD_002
USA_FLO_AD_003
USA_FLO_AD_004
USA_FLO_AD_005
USA_FLO_AD_006
USA_FLO_AD_007
USA_FLO_AD_008
USA_FLO_AD_009
USA_FLO_AD_010
USA_FLO_AD_011
USA_FLO_AD_012
USA_FLO_AD_013
USA_FLO_AD_014
USA_FLO_AD_015
USA_FLO_AD_016
USA_GEO_AD_001
USA_GEO_MF_001
USA_GEO_MF_002
USA_ILL_AD_001
USA_ILL_AD_002
USA_ILL_MF_001
USA_LOU_AD_001
USA_LOU_AD_002
USA_LOU_MF_001
USA_MIP_MF_001
USA_MIS_AD_001
USA_MIS_AD_002
USA_TEX_AD_001
USA_TEX_AD_002
USA_TEX_AD_003
USA_TEX_AD_004
USA_TEX_AD_005
USA_TEX_AD_006
USA_TEX_AD_007
USA_TEX_AD_008
USA_TEX_AD_009
USA_TEX_AD_010
USA_TEX_AD_011
USA_TEX_AD_012
USA_TEX_AD_013
USA_TEX_AD_014
USA_TEX_AD_015
USA_WIS_AD_001
VNM_HAI_AD_001
VNM_HAN_AD_001
Coverage per 100kb window
Genome position
CHR
chr1
chr2
chr3
chr4
chrX

856

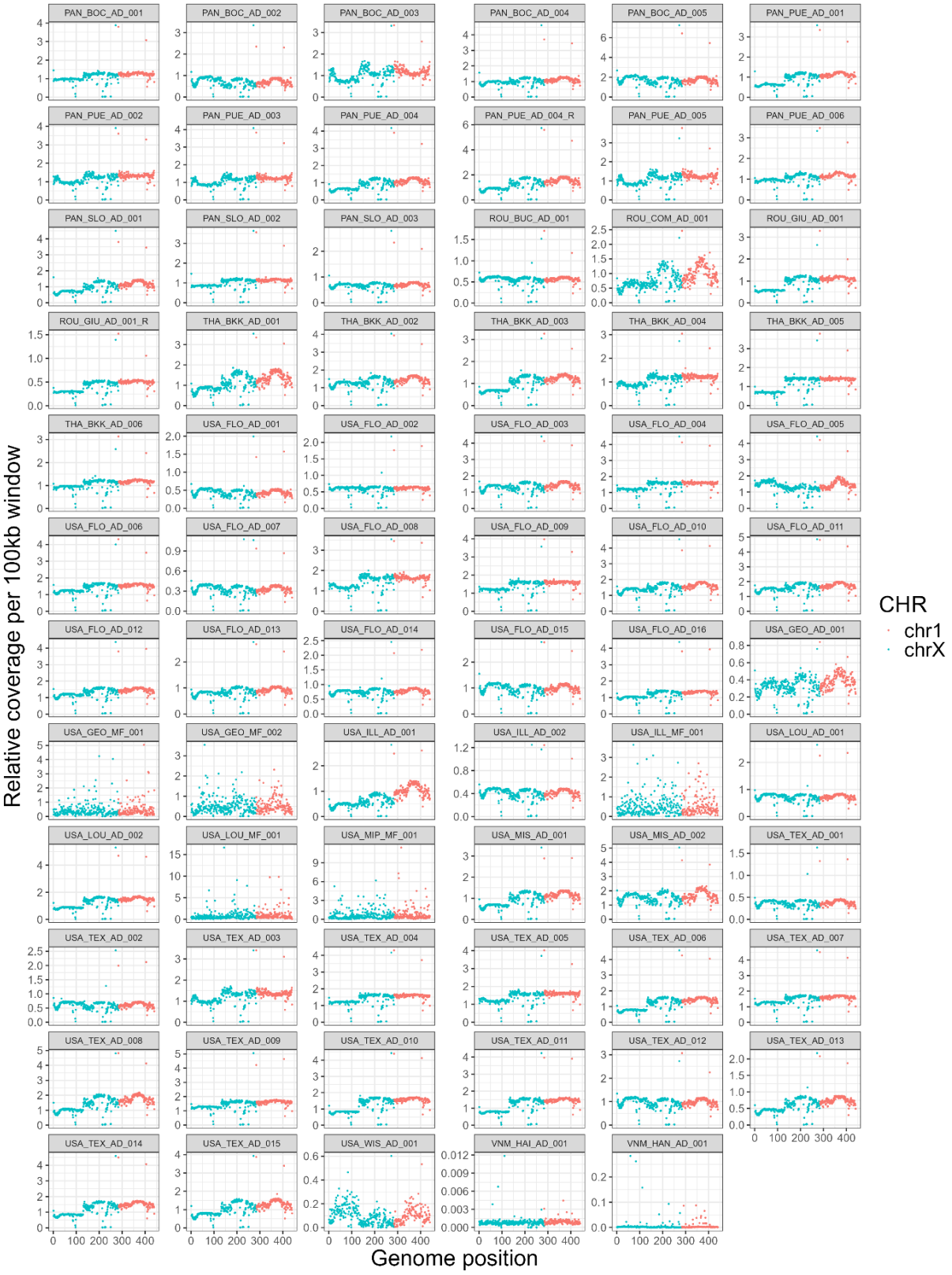
857 **Supplementary Fig. 2: Raw nuclear sequencing coverage per sample (Part 2/2).** Coverage
858 was calculated using 100 kb sliding windows for all mapped samples. Each chromosome is
859 shown in a different colour, with a legend on the right.

860




**Supplementary Fig. 3: Sex determination of samples based on relative coverage of sex-linked vs autosomal chromosomes (Part 1/2).** Sequencing coverage was assessed using a 100kb sliding window and compared to the overall median coverage. Coverage of the sex-

linked X chromosome (blue) was compared to an autosome (chromosome 1, red) to determine the parasite's sex. Male canine heartworms show reduced coverage on chromosome X (ratio ~0.5), while females show similar coverage across sex-linked and autosomal chromosomes (ratio ~1). Ambiguous results were observed in samples with low coverage. The remaining samples are shown in Supplementary Fig. 4.




**Supplementary Fig. 4: Sex determination of samples based on relative coverage of sex-linked vs autosomal chromosomes (Part 2/2).** Sequencing coverage was assessed using a 100kb sliding window and compared to the overall median coverage. Coverage of the sex-

875 linked X chromosome (blue) was compared to an autosome (chromosome 1, red) to determine
876 the parasite's sex. Male canine heartworms show reduced coverage on chromosome X (ratio
877 ~0.5), while females show similar coverage across sex-linked and autosomal chromosomes
878 (ratio ~1). Ambiguous results were observed in samples with low coverage.