Equus roundworms (Parascaris spp.) are undergoing divergence due to natural and anthropogenic factors

Lei Han[1], Tianming Lan[2], Yaxian Lu[1], Mengchao Zhou[1], Haimeng Li[3], Haorong Lu[2], Qing Wang[4], Xiuyun Li[5], Shan Du[6], Chunyu Guan[5], Yong Zhang[7], Sunil Sahu[3], Puyi Qian[3], Shaofang Zhang[3], Hongcheng Zhou[3], Wei Guo[8], Hongliang Chai[1], Sibo Wang[3], Quan Liu[1], Huan Liu[9], and Zhijun Hou[1]

[1]Northeast Forestry University
[2]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen
[3]BGI-Shenzhen
[4]University of Chinese Academy of Sciences
[5]Harbin northern forest zoo
[6]Inner Mongolia Agriculture University
[7]893386379@qq.comCenter for Animal Disease Control and Prevention of Ordos
[8]Chinese Academy of Agricultural Sciences Harbin Veterinary Research Institute
[9]University of Copenhagen

March 07, 2024

Abstract

The evolution of parasites is often directly affected by the host's environment or behavior. Studies on the evolution of the same parasites in different hosts are extremely attractive and highly relevant to our understanding of divergence and speciation. Here we analyzed the genetic variation of Equus roundworm populations in different hosts (horses, zebras and donkeys), and presented the first molecular evidence of divergence in Equus roundworms (Parascaris univalens). At the genetic level, Equus roundworms were mainly separated into two clades (Horse-derived and Zebra & Donkey-derived). This divergence began at 600-1500 years ago, which interestingly coincided with the domestication history of horses. We found that compared with horse-derived roundworms, most of the key enzymes related to glycolysis were under strong positive selection in zebra & donkey-derived roundworms, indicating that the evolution of the metabolic level was one of the main reasons for the divergence. In addition, we conducted a selective scan of resistance-related genes and found that the three populations were under different degrees of selection. This prompted us to pay attention to the possible impact of drugs on divergence, not just the drug resistance. This work supports that divergence or speciation is a continuous and dynamic process, and continuous monitoring of environmental factors is conducive to further understanding the adaptive evolution of roundworms.

Equus roundworms (*Parascaris* spp.) are undergoing divergence due to natural and anthropogenic factors

Lei Han[1,5+], Tianming Lan[2,3+*], Yaxian Lu[1+], Mengchao Zhou[1], Haimeng Li[2,4], Haorong Lu[9,10], Qing Wang[2,4], Xiuyun Li[6], Shan Du[7], Chunyu Guan[6], Yong Zhang[8], Sunil Kumar Sahu[2], Puyi Qian[9,10], Shaofang Zhang[9,10], Hongcheng Zhou[9,10], Wei Guo[11], Hongliang Chai[1,5], Sibo Wang[2], Quan Liu[1*], Huan Liu[2,3*], Zhijun Hou[1,5*]

[1]College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China

[2]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China

[3]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

[4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[5]Key Laboratory of Wildlife Conservation, China State Forestry Administration, Harbin 150040, China

[6]Harbin northern forest zoo, Harbin 150040, China

[7]Inner Mongolia Agriculture University,Hohhot, 010000, China

[8]Center for Animal Disease Control and Prevention of Ordos, Inner Mongolia Ordos 017000, China

[9]Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen 518120, China

[10]China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China

[11]State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin150040, China

[14]Qingdao-Europe Advanced Institute for Life Sciences, Qingdao 266555, China

[15]Foping National Nature Reserve, Hanzhong 723400, China

[+]These authors contributed equally to this work.

[*]Corresponding authors: houzhijundz@163.com, liuhuan@genomics.cn, lantianming@genomics.cn, liuquan1973@hotmail.com

Abstract

The evolution of parasites is often directly affected by the host's environment or behavior. Studies on the evolution of the same parasites in different hosts are extremely attractive and highly relevant to our understanding of divergence and speciation. Here we analyzed the genetic variation of Equus roundworm populations in different hosts (horses, zebras and donkeys), and presented the first molecular evidence of divergence in Equus roundworms (*Parascaris univalens*). At the genetic level, Equus roundworms were mainly separated into two clades (Horse-derived and Zebra & Donkey-derived). This divergence began at 600-1500 years ago, which interestingly coincided with the domestication history of horses. We found that compared with horse-derived roundworms, most of the key enzymes related to glycolysis were under strong positive selection in zebra & donkey-derived roundworms, indicating that the evolution of the metabolic level was one of the main reasons for the divergence. In addition, we conducted a selective scan of resistance-related genes and found that the three populations were under different degrees of selection. This prompted us to pay attention to the possible impact of drugs on divergence, not just the drug resistance. This work supports that divergence or speciation is a continuous and dynamic process, and continuous monitoring of environmental factors is conducive to further understanding the adaptive evolution of roundworms.

Keywords: Diversification, *Parascaris univalens* , adaptation, evolution

1 | INTRODUCTION

The domestic horse (*Equus caballus*) is a culturally, economically, and historically important domesticated animal. The domestication of the horse in central Asia ˜5,000 years ago is one of the most significant achievements of human civilization (Outram et al., 2009). Since then, horse has been extensively used in agriculture, transportation, war, and sports (Kalbfleisch et al., 2018). A significant change of the impact of domestication on animals is mainly in nutrition and behavior, but the impact on the large parasites living in the host remains unknown. The genetic architecture and selection are key indicators of evolution aiming to understand how natural selection and artificial selection can lead to lineage divergence and speciation(Jones et al., 2012). The genetic differences between populations can reveal a lot about the basic evolutionary process due to changes in the environment. (Schluter, 2009). Studying the adaptation of parasites along with host domestication is of great significance for understanding the host-parasite interactions.

Equine roundworms are large parasitic nematodes that predominantly infect foals and weanlings. The *Equus* species, such as horse (*Equus caballus*), zebra (*Equus zebra*) and donkey (*Equus asinus*), are the reservoir hosts for *Parascaris*spp. Despite their close kinship, the extent of domestication in these hosts are vastly different. Horses have been domesticated for thousands of years and are intimately linked to human activity. Comparatively, donkeys are far less domesticated than horses, while zebra has not been domesticated so far. In addition, the nutritional requirements of the three are also different. For example, donkeys have much lower energy and protein requirements than those of other equids(Martin-Rosset, 2018), and even the metabolic response during exercise will differ depending on the level of food given to the horse (Jansson & Lindberg, 2012), which may also be a challenge for parasites. Extensive genetic diversity provides the genetic basis for the adaptive evolution of nematodes, but a successful evolution will also come at a cost to the host (Hay et al., 2017). Roundworms' underlying genetic diversity contributes to their ability to adapt and undergo selective evolution. Understanding these processes, as well as the key environmental factors that lead to selection, is crucial to comprehend the evolution of parasites (Salle et al., 2019).

The control of parasitic nematodes feeding on animal relies almost exclusively on anthelmintic drugs, which is proved to be effective in the short-term management, but the long-term effectiveness has been questioned due to the widespread emergence of drug resistance (Kaplan & Vidyashankar, 2012). There is widespread concern about the risk associated with relying on anthelmintics with hundreds of millions of doses being donated and used every year(McManus, 2018). In general, levels and spectrum of anthelmintic resistance are less severe in parasites of horses, however, the same issues persist, and they seem to be worsening (Kaplan, 2004; Kaplan et al., 2004; Traversa et al., 2009). It is now clear that the effective method for long-term control of nematodes is to improve pasture (or nature reserve) management and scientific genetic monitoring, rather than simply using chemical methods (Kaplan & Vidyashankar, 2012). In recent years, genome scanning has become an effective means to reveal the genetic determinants of habitat differences in some organisms. Since less introgression occurs in the selected loci, they exhibit lower polymorphism than other regions of the genome, which enables the formation of highly divergent regions that serve as the genetic foundation for divergence (Nosil, Vines, & Funk, 2005; Wu, 2001). Here we analyzed the genome characteristics of*Parascaris* spp. populations dwelling on three main hosts: horse (*Equus caballus*), zebra (*Equus zebra*) and donkey (*Equus asinus*). To the best of our knowledge, here we present the first report on the recent divergence of *Parascaris* spp. populations, and speculate that it may be linked to roundworm host preference. Meanwhile, the role of selection (natural or man-made selection) in the population was also evaluated at the genetic level. This work is of great significance for changing the current "one size fits all" roundworm taxonomy, and opens a new era of precise monitoring of roundworm evolution.

2 | MATERIALS AND METHODS

2.1 | Samples

Seventeen roundworm individuals were collected from the naturally infected horses (*Equus caballus*) treated with anthelmintics in a farm located in the Ordos, Inner Mongolia, China. While, two horse roundworm individuals were collected after anthelmintic treatment from a farm in Harbin, Heilongjiang, China. Five donkey (*Equus asinus*) roundworm individuals were collected after anthelmintic treatment from a farm in Chifeng, Inner Mongolia, China. Eighteen roundworms from zebra (*Equus zebra*) were obtained from Harbin Northern Forest Zoo, Heilongjiang, China. All specimens were washed extensively in sterile physiological saline (37), snap-frozen and transported with dry ice and then stored at -80 until further use. All experimental designs and nematode handling were approved by the Institutional Animal Care and Use Committee of Northeast Forestry University.

2.2 | Nucleic acid isolation, library construction and sequencing

Total genomic DNA was isolated using a sodium dodecyl sulphate/proteinase K digestion (Gasser et al., 2006) followed by phenol-chloroform extraction and ethanol precipitation. Genomic DNA was sheared into 200-800 bp for paired-end libraries preparation according to the manufacturer's instructions of the DIPSEQ platform (BGI-Shenzhen, Shenzhen, China). Libraries were then subjected to the DIPSEQ-T1 sequencer for

short-read whole genome sequencing (WGS) sequencing (Table S1).

2.3 | Read mapping and SNP calling

High-quality reads were aligned to the *P. univalens* reference genome (WormBase accession ID: GCA_002259215.1) using BWA-MEM (0.7.13-r1126)(H. Li & Durbin, 2010) with default parameters. SAMtools (v0.1.19) (H. Li et al., 2009) was used to convert mapping results into the BAM format and filtered the unmapped and non-unique mapping reads. The *Parascaris equorum* reference genome (WormBase project ID: PRJEB514) was used for species identification. Duplicated reads were marked with the Picard Tools (picard.sourceforge.net, Version: 2.1.1). Then Genome Analysis Toolkit (GATK v 4.0.3.0) (Depristo, Banks, Poplin, Garimella, & Daly, 2011) were used to population SNP calling. Then hard filtering was applied to the raw variant set using "QD < 2.0 || FS > 60.0 || MQ < 20.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" –filter-name "snp_filter". SNPs with >1% missing data or <0.01 minor allele frequency (MAF) were filtered out using vcftools (v0.1.12a) (Danecek et al., 2011) for downstream bioinformatic analyses. Variants Annotation of variantswere performed according to the reference genome using the package ANNOVAR (Version: 2015-12-14). Using the SAMtools software, the coverage of each sample was counted based on aligned BAM data.

2.4 | Diversity analysis

SNP density was counted with a 10 kb sliding window using VCFtools (v0.1.13) software (Danecek et al., 2011). Genome-wide nucleotide diversity (π) and Tajima's D were computed by sliding windows of 1 kb using all individuals in each population using VCFtools (v0.1.13). The Weir-Cockerham fixation index (*Fst*) was estimated among three populations with genotype data using the VCFtools (v0.1.13).

2.5 | Phylogenetic relationship, genetic structure and admixture

Principal component analysis (PCA) was carried out using EIGENSOFT (Nick et al., 2006) software base on the SNP dataset, and the population clustering analysis was conducted in PLINK (v1.90b6.10) (Purcell et al., 2007). We used genome-wide SNPs to construct the maximum likelihood (ML)phylogenetic tree with 1000 bootstraps using iqtree (v1.6.12) (Lam-Tung, Schmidt, Arndt, & Quang, 2015). The genome sequence of*Baylisascaris schroederi* was selected as an outgroup. Population structure was analyzed using the ADMIXTURE (v1.3.0) program with a block-relaxation algorithm. To explore the convergence of individuals, we predefined the number of genetic clusters of K from 2 to 4.

We investigated the relationships within *Parascaris* spp. populations in a coalescent framework with SNAPP implemented in BEAST v2.6.3 (Bouckaert et al., 2014). We performed two independent runs with a chain length of 10,000,000 generations, sampling every 1,000 generations. We examined convergence using TRACER v1.7.1 and created a maximum clade credibility tree after a burn-in of 20% via TREEANNOTATOR (Helfrich, Rieb, Abrami, Lücking, & Mehler, 2018). According to epidemiological investigation, we assumed that the average generation time of *Parascaris* spp. was 0.17 years, and converted the SNAPP analyses into units of real-time using a mutation rate (μ) of 9×10^{-9} per generation per site.

2.6 | Estimates of the effective population size and divergence time

The pairwise sequentially Markovian coalescent (PSMC) method (Heng Li & Durbin, 2011) was used to evaluate the dynamic change of effective population size (*Ne*) of each population. We used 0.17 years per generation (g) and mutation rate (μ) of 9×10^{-9} per generation per site to rescale the time to year (Cutter, 2008). More recent (within 1000 years) changes of effective population size of each population and separation time between different populations were further estimated by using the MSMC2 (Schiffels & Durbin, 2014), which can much compensate results from PSMC. However, the inference accuracy of MSMC2 largely depends on the phasing accuracy of genotypes. Switch error rates will introduce bias in the calculation. To further confirm results from multiple sequentially Markovian coalescent (MSMC2), we also used Sequentially Markovian Coalescent (SMC++) methods (Terhorst, Kamm, & Song, 2017) to do the same analysis as of MSMC2. The SMC++ used phasing-free genotype data to do the population history and separation time inference, which becomes a reliable method to support inferences from MSMC2. For

SMC++, we set the upper bound for the number of generations to 10,000 to estimate size history, and calculate the lower bound based on a heuristic approach. For MSMC2, we first phased all SNPs of each individual by using beagle (v5.0) (Browning & Browning, 2007), then the calculation was performed with the following parameters: -i 20 -t 6 -p '10*1+15*2'. The mutation rate (μ) of *Parascaris* spp. for SMC++ and MSMC2 were used the same values as for PSMC.

2.7 | Δεμογραπηις ινφερενζε υσινγ φαστσιμζοαλ2 ανδ δαδι

We used the fastsimcoal2 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013) approach to deduce the recent demographic history of*Parascaris* spp. populations. We chose only SNPs located in intergenic regions to avoid the influence of SNPs under selection(Zhou et al., 2018). We used δaδi (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) to investigate alternative demographic scenarios for the species complex. In the absence of historical evidence, we hypothesized that there may or may not be any form of gene flow between roundworm populations. In order to get the best model, we first simulated a total of ten δaδi 3D models, including one simple model, three simultaneous splitting models, one ancient migration model, one simultaneous splitting variations model, one admixed ("Hybrid") origin model and three divergences with gene flow variations models. When fitting demographic models, we perform multiple runs (100 rounds) and ensure that final optimizations are converging on a similar log-likelihood score. To estimate demographic parameters, the derivative-based BFGS algorithm was used to optimize the composite log-likelihood. All models and scripts are available at*https://github.com/dportik/dadi_pipeline*.

2.8 | Recent nature selection analysis

Extended Haplotype Homozygosity (EHH) and Integrated Haplotype Score (iHS) methods were used for detecting SNPs under recently positive selection of three roundworm populations (Mathieu & Renaud, 2012). We use SNPs with an iHS score of top 0.5% and the distance between adjacent SNPs $<$ 50 kb as candidate SNPs (Akagi, Hanada, Yaegaki, Gradziel, & Tao, 2016). We searched for genes in the 5-kb flanking region from both sides of candidate SNPs, and calculate the accumulated iHS scores by adding all iHS scores of candidate genes. Next, to uncover genetic variants under strong positive selection in each host population, we used XP-extended haplotype homozygosity (XP-EHH) method on each pair of combination (horse roundworm (PEc) vs zebra roundworm (PEz), horse roundworm (PEc) vs donkey roundworm (PEa) and PEz vs PEa) to find population-specific SNPs under strong positive selection. XP-EHH we used in this study was from the R package rehh (v3.1.2; https://cran.r-project.org/web/packages/rehh/vignettes/rehh.html). The regions with P values $<$ 0.01 were considered significant signals in the population of interest.

3 | RESULTS

3.1 | Genome resequencing and genetic variation

A total of 44 individuals from three*Parascaris* spp. populations of horses, zebra, donkey from Inner Mongolia and Heilongjiang of China were whole-genome sequenced (Fig. 1). We identified 4,398,519 SNPs with a genome-wide distribution of 1 SNP per 57 bp on average (Fig. S1). Genome resequencing was accomplished with an average depth of 20X (Fig. S2), average mapping rate of 98.17% and with $>$ 90% genomic coverage of all individuals (Table S2). Currently there is a dispute over whether there are two types of horse roundworms (*P. equorum* or*P. univalens*) or not. The BWA alignment confirmed that the all individuals were closer to *P. univalens* , with an average mapping rate against *P. equorum* of 78%, significantly lower than that against the *P. univalens*genome (P $<$0.01) (Fig. S3a). We have counted the number of shared SNPs among the three populations, and found that 1,244,910 SNPs were shared among them. PEc has 1,055,384 unique SNPs, while PEz and PEa have only 206,419 and 65,110 respectively. This also showed the genetic difference between PEc and the other two populations. In addition, we calculated the observed heterozygosity and expected heterozygosity of individuals in populations to ensure that each population is represented (Table S3).

3.2 | Genetic structure and phylogenetic relationships

Principal components analysis (PCA) supported the clear separation among*Parascaris* spp. (Fig. 2a), with

PC1 and PC2 separating the PEc and PEz&PEa populations (P <0.05). All PCs showed that PEz and PEa were located in the same cluster. Meanwhile, the phylogenetic relationship among the three populations inferred by ML tree highlighted a similar division as that of PCA (Fig. 2b). The tree showed two distinct clusters, the PEc population seems to be a separate clade, while the other two formed a distinct clade. Although the sampling sites of PEc and PEz partially overlapped, the divergence between them was still clear. Population structure analysis further confirmed the two distinct clusters presented by PCA and ML tree, where PEz and PEa had more ancestral components in common (Fig. 2c). We scanned the paired IBD regions at the genome-wide level of all individuals and found that PEz and PEa shared more IBD regions (98.5% of PEa shared with PEz), while PEc and the other two populations had almost no shared large fragments (Fig. S4, S5). In addition, lower inter-population *Fst* values were detected in PEz and PEa populations also indicating a closer relationship between them (Fig. S5).

3.3 | Demographic history and divergence time

To examine the genome-wide divergence time among *Parascaris* spp. populations on the genome-wide, we constructed a molecular clock phylogenetic tree based on the calibrated mutation rate using all SNP sites. The tree topology showed that the divergence time of PEc and PEz&PEa was about 900-1500 years ago when the posterior probability was >95% (Fig. 3). The PSMC results showed similar effective population sizes (*Ne*) of three populations (Fig. 4a). Besides, the MSMC2 inferred that the *Ne* of the three populations had almost a similar trend 1000 years before, without any differentiation, indicating the existence of possibly common ancestor (Fig. 4a, 4b). However, the *Ne* of the three populations began to diverge in the recent 1000 years (Fig 4b). The relative cross coalescence rates (RCCR) estimation between pairs of the three populations also showed that clear genetic separations occurred from 600 to 1500 years ago, with an earlier separation of PEc with PEz and PEa than that of PEz and PEa (Figure 4b). The change of *Ne* of each population and separation among the three population were further supported and validated by results inferred by SMC++ (Figure 4c). We carefully compared the relationship between the separation and the topological structure of the phylogenetic tree, and found high consistency. Both results showed that PEc vs PEz, PEc vs PEa had obvious divergence, but PEz vs PEa are not fully differentiated. Taken together, we considered that *Parascaris*spp. mainly produce two divergence clades, one from horse-derived (PEc) and the other from zebra & donkey-derived (PEz&PEa). The conservative divergence time was estimated to be about 200-1500 years ago. Finally, we used fastsimcoal2 to evaluate the population size after divergence based on the observed joint site frequency spectrum (SFS; Fig. S6), and found that the PEc population size after divergence event was significantly larger than PEz&PEa (Fig. 4d). The best coalescent simulation model inferred by fastsimcoal2 also indicates an early bidirectional gene flow between PEc and PEz&PEa.

3.4 | Genetic changes and migration events

We used the δaδi to further explore the demographic history prior to the separation of the three populations. Following the method of Portik (Portik et al., 2017), We employed a four-round optimization technique to ensure that all final optimizations resulted in a similar log-likelihood score (Table S4). In order to better validate the possible divergence patterns between populations, we first constructed ten 3D models for three independent populations (Fig. S7). Models with the lowest scoring log-likelihoods was “Adjacent ancient migration, shorter isolation” (Fig. 5a, Table S5). This ancient migration model involving early gene flow with symmetric migration was supported as the best fit for the three populations. Next, in our eight 2D simulations (Fig. S8), the ancient migration or secondary contact plus instantaneous size change model involving divergence with ancient continuous symmetrical migration, isolation with instantaneous size change provided the best fit regarding the PEc and PEz&PEa lineages (Fig. 5b, Table S5). By comparing the best 2D and 3D models, it was found that the best 2D models have larger log values and lower residuals, which coincided with the results of our phylogenetic tree. Furthermore, the best model revealed a possible divergence of roundworm populations, that is, there was bidirectional gene flow in the early ancient periods, but in the middle and current stages, the gene flow between the populations almost ceased.

3.5 | Adaptive evolution in metabolism

In-depth genome scanning and functional annotation helped us to understand the divergence event. Genome-wide nucleotide diversity (π) was computed for each population with all individuals. Meanwhile, we identified genomic regions as candidate divergent regions (CDRs) among PEc, PEz and PEa populations (Table S6, Fig. S9). We used iHS to detect genes under recent natural selection in the PEc and PEz&PEa populations. A total of 1,046 SNPs in PEc and 1,093 SNPs in PEz&PEa were identified within the top 1% iHS scores (Fig. S10). These positively selected significant SNPs were annotated to 290 and 254 functional genes, respectively. The GO functional enrichment showed that they were mainly enriched in GO terms such as metabolism and regulation of gene expression (Table S7). The results of KEGG enrichment also showed that the two clades have significant selection signals in metabolic-related signaling pathways (Fig. S11). In addition, we also used the XP-EHH method to screen for genes that may have been positively selected by different environmental pressures by comparing the PEc and PEz&PEa populations. The two-side P -value test was used to scan genome regions with selection sweep signals of the two clades. Interestingly, the differences in carbohydrate metabolism and lipid metabolism were extremely significant in the two clades (Fig. S12 – S13, Table S8). PEz&PEa clade has shown significant positive selection for almost all key enzymes in glycolysis and tricarboxylic acid cycle. These loci have aroused our attention and re-examined the selection dynamics of their surrounding regions (Fig. 6). This includes the kinases (E1:hexokinase and E2:6-phosphofructokinase-1) involved in the two most important irreversible reactions in the first stage of the conversion of glucose to pyruvate under anaerobic conditions. Meanwhile, the dehydrogenase (isocitrate dehydrogenase) in the irreversible reaction of isocitrate oxidative decarboxylation to α-ketoglutarate has also been significantly positively selected (P <0.05). The collective selection of enzymes in the glycolysis process and tricarboxylic acid cycle showed that PEz&PEa has a greater demand than PEc in this process. In addition, members of the lipid synthase family which is involved in the uptake of fatty acids are significantly positively selected in PEc. Parasitic helminths contain appreciable quantities of lipids. However, most of the intestinal helminths do not utilize significant amounts of lipids even during starvation and under aerobic conditions(Frayha & Smyth, 1983) largely due to their anaerobic mode of life. The significant selection of these enzymes such as fatty acid CoA synthetase family and long-chain fatty acid CoA ligase 5 suggests that they might be involved in some other processes as well, not just lipid uptake or metabolism. In terms of energy conservation, considering lipids and end products of carbohydrate metabolism for helminths seem unrealistic and uneconomical.

3.6 | Potential drivers of divergence

Currently, the anthelmintics mainly have two modes of action, one is more rapid action on membrane ion channels, and the other is a relatively slow biochemical reaction. These common anthelmintics include benzimidazoles (BZs), macrolides (MLs), nicotinic acetylcholine receptor agonists (Wolf et al.) and aminoacetonitrile derivatives (AADs). We screened out the main genes that may be related to the resistance of all the above-mentioned anthelmintics that have been reported so far (Table 1). We performed selective scanning (iHS) within 50 kb of all these gene regions, and calculated the nucleotide diversity (π) and*tajima'D* of the three populations with 10 kb sliding windows. The results showed that multiple resistance-related genes were strongly selected among different populations (positive or negative; Table 1). For example, we first identified three classic resistance locus (167/Phe, 198/Glu and 200/Phe) of the BZs resistance gene β-tubulin (Lake, Matthews, Kaplan, & Hodgkinson, 2009), and found that all individuals in the three populations showed non-resistant mutations through sequence alignment (Fig. S14a). We then further scanned the selection of these gene regions and found that although there were no drug-resistant mutations in these regions, and they showed significant positive selection (Fig. S14c). In addition, the selection of other resistance-related genes in the population have also been discovered (Fig. S15 - S17). For example, multidrug resistance protein *pgp-3*and multidrug resistance protein 1 (*mrp-1*) revealed strong positive selections in the PEc population. Glutamate-gated chloride channel alpha (*glc-1*), which was related to ivermectin resistance, showed strong positive selection in PEz and PEa populations. The selection of these regions do not have a certain rule, which may be related to the medication history in different environments. However, it was obvious that in some groups, certain genetic selections have been permanent and would continue to be passed down to the next generation.

4 | DISCUSSION

The ecological environment in which parasites inhabit is different from that of ordinary animals, and the survival of parasites is more dependent on the intestinal environment. The survival and evolution of the equine are inextricably linked to nature and human activities. Our study used a combination of explicit genetic analysis and demographic models to determine the possibly diversified mechanisms that occur in different intestinal environments. We applied δaδi 2D/3D models to roundworm populations distributed in different regions and different hosts to unveil whether the diversity of the population was caused by geographic distribution and ecological environment, or simply caused by host specificity, or was related to human domestication of its host. The ancient migration or secondary contact model, as well as the immediate size change model, were found to be effective in explaining the demographic differences and recent divergence of *Parascaris* spp. populations. In addition, the demographic history showed that the*Parascaris* spp. were in the process of divergence. Selection analysis provided evidence for understanding the possible causes of the divergence, which supports the significant impact of different intestinal habitats on evolution. We summarized the main findings on the diversification of *Equus* roundworms, and provided a perspective for future monitoring of roundworm ecology in a timely manner to deal with possible unfavorable mutations.

Host preference as the primary cause of divergence

We have roundworm samples from various hosts from multiple locations around northern China, including grassland pastures, zoos, and family farms. Our clustering analysis revealed that even if a sampling site of the horse roundworm (PEc) was closer to the zebra roundworm (PEz), the horse roundworm still clustered into one clade in terms of genetic relationship. This specificity deserves our attention. It suggests that the single-origin *Equus* roundworm may have differentiated, and this divergence was probably related to the parasitic host. In addition, clustering results also showed that even if they were geographically closer, the host's impact on roundworms was obviously more critical. The host provides habitat and nutrition for the parasites. Apparently, the parasites' local adaptation to the host causes the parasites of the same host to evolve in the same direction (Kaltz & Shykoff, 1998). Studies have shown that a parasite population has higher mean performance in local hosts compared to foreign host populations(Lively, 1996; Schulte, Makus, Hasert, Michiels, & Schulenburg, 2011). When combined with population structure, it is clear that the association between geographical cause and *Parascaris* spp. evolution was not strong.

The impact of nature and domestication on roundworms

The domestication of horses has completely changed the behaviors of warfare and transportation in human society. The usage of horses as the main transport by the Cavalry army appeared in the early Iron Age (Drews, 2004). Studies have shown that the decline in heterozygosity indicates that in recent centuries, the breeding population of horses has been greatly reduced, and this was the inevitable result of domestication and reproduction (Fages et al., 2019). During the war, humans significantly preferred the speed traits of horses. Therefore, in order to maintain the excellent traits of horses, on the one hand, people began to select excellent stallions in breeding (Fages et al., 2019), and on the other hand, the forage nutrition of horses has been gradually improved. The change of the host's diet was an opportunity and a challenge for the parasites. From the perspective of glycolysis, we found that most of the key enzymes involved in glycolysis in zebra roundworms and donkey roundworms were subjected to a higher degree of recent selection when compared to domestic horses. As the most important way for roundworms to obtain ATP, glycolysis and tricarboxylic acid cycle seem to have "degraded" in domestic horse roundworm populations. This may be the result of its adaptation to host domestication. Experiments have shown that the content of fatty acids such as palmitic acid, palmitoleic acid, stearic acid and oleic acid of the parasite were almost the same as that of the specific host and the changes in the ratio of fatty acids tend to be synchronized(Barlow, 1972). Immunological evasion could be the major purpose. This notion is supported by our findings in horse roundworms, which showed a strong selection of genes involved in lipid synthesis. Obviously, in addition to nutrition and maintaining physiological integrity, a more important purpose was possible to keep consistent with the host's various fatty acid patterns. The evolution of parasites in regulating lipid composition may be a factor influencing host suitability (Wallis, 1982). The difference was that we believed the consistency

in this case was more likely related to the host's intestinal and surrounding lipid deposition rather than the total lipid ratio.

The consequences of anthelmintics should not be defined only in resistance

The issue of drug resistance has been widely mentioned over the past two decades. Parasites have strong adaptability, and roundworms lay more than 200,000 eggs per day (Wang & Davis, 2020), which is destined to have a sufficient mutational basis to resist any environmental changes. The problem of drug resistance has brought significant economic losses to industries such as animal husbandry, and the control of parasites has also become an important expenditure (McKellar & Jackson, 2004). We found that some resistance-related genes such as *β-τυβυλιν* ,*glc-1* , *pgp-3* , *mrp-6* , *cup-4* , *nrf-6* and CYP family, were significantly selected in different populations. These genes are respectively related to multiple anthelmintics, which may be related to their history of deworming. Over time, the gene frequency of these genes will increase significantly in the population and attention should be paid to the impact of this selection on species evolution. Current vaccines offer an attractive alternate control strategy against these parasites (Hewitson & Maizels, 2014), and also require a huge investment in early-stage research and development. Despite the wealth of methods, the response of the parasites was amazing. When we focus on the issue of drug resistance, we should worry more about the impact of this strong selection on species evolution. Our results indicated the influence of human activities on the evolutionary selection of*Equus* roundworms, and this influence brings irreversible parasitic preference. Human anthelmintic medications are nowadays chosen in a more direct manner than the host's domestication, and the frequency of use is more frequent. However, the evolutionary consequences of short-term multi-generation strong selection are still uncertain. This suggests that we should be cautious in dealing with the issue of drug resistance and adopt more scientific strategies.

5 | CONCLUSIONS

Roundworms has undergone significant differentiation and host specificity in the process of accompanying host evolution. This can be clearly seen in the evidence in morphology and molecular biology. The roundworm genome provides the possibility to study the details of gene selection or loss in the process of roundworm differentiation. It is also the case that some of the genes we have identified have potentially allowed new avenues for gene selection and intestinal environment adaption, for example, epidermal chitin synthesis, environmental information and essential amino acid metabolism. In addition, population genomics analysis and drug prediction provide insights for revealing the impact of deworming history on population genetic structure and prevention.

ACKNOWLEDGMENT

This work was supported by the National Key R&D Program (No. 2017YFD0501702), Open Project of Key Laboratory of SFGA on Conservation Biology of Rare Animals in The Giant Panda National Park (CCRCGP, No.2020004), Forestry science and technology research project (No. 20180302), the Pearl River Talent Recruitment Program in Guangdong Province (2019CX01N111), Fundamental Research Funds for the Central Universities (No. 2572020AA30) of China, the Foundation of Key Laboratory of State Forestry and Grassland Administration (State Park Administration) on Conservation Biology of Rare Animals in the Giant Panda National Park (No. KLSFGAGP2020.002), and the Guangdong Provincial Key Laboratory of Genome Read and Write (grant No. 2017B030301011).

AUTHOR CONTRIBUTION

Z.J.H., H.L. and T.M.L. designed and initiated the project. Y.X.L., X.Y.L., S.D. G.C.Y., Y.Z., and L.H. collected the samples. L.H., S.F.Z., Q.P.Y., H.C.Z., and H.R.L. performed the DNA extraction, library construction and sequencing. L.H., T.M.L., H.M.L., and Q.W. performed the data analysis. L.H., and T.M.L. wrote the manuscript. Z.J.H., H.L., T.M.L., Q.L., S.B.W., W.G., S.K.S., and Y.X.L. supervised the manuscript. All authors read and approved the final manuscript. L.H. and T.M.L. wrote the manuscript with input from Z.J.H., H.L., T.M.L., Q.L., H.C.Z., S.B.W., S.K.S., and Y.X.L. supervised the project.

CONFLICT OF INTERESTS

The authors declare no conflict financial interests.

DATA AVAILABILITY

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) (Guo et al., 2020) of China National GeneBank DataBase (CNGBdb) (Feng, Li, Fan, Li, & Xiao, 2020) with accession number CNP0001875.

REFERENCES

Akagi, T., Hanada, T., Yaegaki, H., Gradziel, T. M., & Tao, R. (2016). Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA research, 23* (3), 271-282.

Barlow, J. (1972). Some host-parasite relationships in fatty acid metabolism. *Insect and Mite Nutrition'(ed. by Rodriguez, JG)* , 437-453.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., . . . Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol, 10* (4), e1003537.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics, 81* (5), 1084-1097.

Cutter, A. D. (2008). Divergence Times in Caenorhabditis and Drosophila Inferred from Direct Estimates of the Neutral Mutation Rate.*Molecular Biology & Evolution* (4), 4.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27* (15), 2156-2158.

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics, 43* (5), 491-498.

Drews, R. (2004). *Early riders: The beginnings of mounted warfare in Asia and Europe* : Routledge.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data.*PLoS genet, 9* (10), e1003905.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., . . . Albizuri, S. (2019). Tracking five millennia of horse management with extensive ancient genome time series. *Cell, 177* (6), 1419-1435. e1431.

Feng, Z. C., Li, J. Y., Fan, Y., Li, N. W., & Xiao, F. W. (2020). CNGBdb: China National GeneBank DataBase. *Hereditas, 42* (8), 799-809.

Frayha, G., & Smyth, J. (1983). Lipid metabolism in parasitic helminths. *Advances in Parasitology, 22* , 309-387.

Gasser, R. B., Hu, M., Chilton, N. B., Campbell, B. E., Jex, A. J., Otranto, D., . . . Zhu, X. (2006). Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nature Protocols, 1* (6), 3121-3128.

Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., . . . Peng, C. (2020). CNSA: a data repository for archiving omics data.*Database, 2020* .

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genet, 5* (10), e1000695.

Hay, S. I., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., . . . Abera, S. F. (2017). Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries

and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The lancet, 390* (10100), 1260-1344.

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., & Mehler, A. (2018). *TreeAnnotator: versatile visual annotation of hierarchical text relations.* Paper presented at the Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).

Hewitson, J. P., & Maizels, R. M. (2014). Vaccination against helminth parasite infections. *Expert review of vaccines, 13* (4), 473-487.

Jansson, A., & Lindberg, J. E. (2012). A forage-only diet alters the metabolic response of horses in training. *Animal, 6* (12), 1939-1946.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., . . . White, S. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature, 484* (7392), 55-61.

Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., . . . Saremi, N. F. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications biology, 1* (1), 1-8.

Kaltz, O., & Shykoff, J. A. (1998). Local adaptation in host–parasite systems. *Heredity, 81* (4), 361-370.

Kaplan, R. M. (2004). Drug resistance in nematodes of veterinary importance: a status report. *Trends in parasitology, 20* (10), 477-481.

Kaplan, R. M., Klei, T. R., Lyons, E. T., Lester, G., Courtney, C. H., French, D. D., . . . Zhao, Y. (2004). Prevalence of anthelmintic resistant cyathostomes on horse farms. *Journal of the American Veterinary Medical Association, 225* (6), 903-910.

Kaplan, R. M., & Vidyashankar, A. N. (2012). An inconvenient truth: global worming and anthelmintic resistance. *Veterinary Parasitology, 186* (1-2), 70-78.

Lake, S. L., Matthews, J. B., Kaplan, R. M., & Hodgkinson, J. E. (2009). *Determination of genomic DNA sequences for beta-tubulin isotype 1 from multiple species of cyathostomin and detection of resistance alleles in third-stage larvae from horses with naturally acquired infections.* Paper presented at the Parasites & vectors.

Lam-Tung, N., Schmidt, H. A., Arndt, V. H., & Quang, M. B. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology & Evolution* (1), 268-274.

Li, H., & Durbin, R. (2010). Fast and accurate short read alignment with Burrows-Wheeler transform.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475* (7357), p.493-496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25* (16), 2078-2079.

Lively, C. M. (1996). Host-parasite coevolution and sex. *Bioscience, 46* (2), 107-114.

Martin-Rosset, W. (2018). Donkey nutrition and feeding: Nutrient requirements and recommended allowances—A review and prospect. *Journal of equine veterinary science, 65* , 75-85.

Mathieu, G., & Renaud, V. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* (8), 1176-1177.

McKellar, Q. A., & Jackson, F. (2004). Veterinary anthelmintics: old and new. *Trends in parasitology, 20* (10), 456-461.

McManus, D. (2018). Schistosomiasis. Nature reviews. Disease primers 4, 13. In.

Nick, Patterson, Alkes, L., Price, David, & Reich. (2006). Population structure and eigenanalysis. *Plos Genetics* .

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats.*Evolution, 59* (4), 705-719.

Outram, A. K., Stear, N. A., Bendrey, R., Olsen, S., Kasparov, A., Zaibert, V., . . . Evershed, R. P. (2009). The earliest horse harnessing and milking. *Science, 323* (5919), 1332-1335.

Portik, D. M., Leaché, A. D., Rivera, D., Barej, M. F., Burger, M., Hirschfeld, M., . . . Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular ecology, 26* (19), 5245-5263.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81* (3), 559-575.

Salle, G., Doyle, S. R., Cortet, J., Cabaret, J., Berriman, M., Holroyd, N., & Cotton, J. A. (2019). The global diversity of Haemonchus contortus is shaped by human intervention and climate. *Nature communications, 10* (1), 1-14.

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics, 46* (8), 919-925.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science, 323* (5915), 737-741.

Schulte, R. D., Makus, C., Hasert, B., Michiels, N. K., & Schulenburg, H. (2011). Host–parasite local adaptation after experimental coevolution of Caenorhabditis elegans and its microparasite Bacillus thuringiensis. *Proceedings of the Royal Society B: Biological Sciences, 278* (1719), 2832-2839.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes.*Nature Genetics, 49* (2), 303-309.

Traversa, D., von Samson-Himmelstjerna, G., Demeler, J., Milillo, P., Schurmann, S., Barnes, H., . . . Beraldo, P. (2009). *Anthelmintic resistance in cyathostomin populations from horse yards in Italy, United Kingdom and Germany.* Paper presented at the Parasites & vectors.

Wallis, R. C. (1982). Current Topics in Insect Endocrinology and Nutrition. *The Yale journal of biology and medicine, 54* (6), 503-504.

Wang, J., & Davis, R. E. (2020). Ascaris. *Current biology: CB, 30* (10), R423-R425.

Wolf, P. G., Rowe, C. A., Kinosian, S. P., Der, J. P., Lockhart, P. J., Shepherd, L. D., . . . Thomson, J. A. (2019). Worldwide relationships in the fern genus Pteridium (bracken) based on nuclear genome markers.*Am J Bot, 106* (10), 1365-1376. doi:10.1002/ajb2.1365

Wu, C. I. (2001). The genic view of the process of speciation.*Journal of evolutionary biology, 14* (6), 851-865.

Zhou, Z., Li, M., Cheng, H., Fan, W., Yuan, Z., Gao, Q., . . . Hu, J. (2018). An intercross population study reveals genes associated with body size and plumage color in ducks. *Nature communications, 9* (1), 1-10.

Tables

Table 1. Summary of the features of the*B. schroederi* genome.

Scaffold ID	Start	End	Description[1]	π	π	π	*Tajima'D*	*Tajima'D*	*Tajima'D*
				PEc^2	PEz^3	PEa^4	PEc^2	PEz^3	PEa^4
PgR045	1372050	1381671	β tubulin	0.0017	0.0019	0.0005	1.4698	2.5324	1.268
PgR004	1320231	1320527	*glc-1*	0.0021	0.0017	0.0007	-0.4035	2.0857	-1.8797

Scaffold ID	Start	End	Description[1]	π	π	π	*Tajima'D*	*Tajima'D*	*Tajima'D*
PgB04	2029204	2057834	*pgp-1*	0.0099	0.0095	0.008	1.5276	1.3328	0.0662
PgR063X	121395	124033	*pgp-3*	0.0002	0.0007	0.0002	0.3278	-0.0274	-1.8733
PgB05	571021	589065	*mrp-7*	0.0119	0.0089	0.0097	-0.2053	1.2363	0.4773
PgR035X	1170116	1202939	*mrp-6*	0.0051	0.0021	0.0047	2.0705	0.7746	0.6784
PgB09	815741	828043	*mrp-5*	0.0053	0.0037	0.002	0.1012	0.6472	-0.6395
PgB10	552422	565520	*mrp-4*	0.0093	0.0099	0.0092	0.9641	1.3762	1.5909
PgB03	2066848	2079134	*haf-2*	0.0033	0.0036	0.0047	0.783	1.4183	0.811
PgR004	2938177	2953151	*haf-5*	0.0232	0.02	0.0173	0.7172	1.6294	-0.3418
PgR006	879399	911228	*cup-4*	0.0083	0.0079	0.0075	2.0797	2.1662	0.6134
PgR009X	834742	840950	*bre-4*	0.0064	0.002	0.0004	1.2283	1.3452	0.7223
PgR004	1575226	1593120	*unc-29*	0.0031	0.0046	0.0036	0.1117	1.9462	1.3307
PgB12X	400372	407591	*unc-38*	0.0003	0.0001	0.0004	-1.4946	0.0306	-0.2792
PgR024	1427386	1436555	Glycosyl hydrolase 18	0.0049	0.0074	0.002	-0.7514	0.6444	0.2132
PgR003	3165465	3176989	nrf-6	0.0052	0.0051	0.0034	2.4611	3.49	2.2507
PgR008	3213688	3247697	ced-7	0.0082	0.0067	0.0064	0.5927	0.2795	0.3853
PgR001X	1705437	1736423	aex-3	0.0018	0.0005	0.0011	0.2075	2.0238	-0.1421
PgB07	1545397	1551068	CYP_2B19	0.0013	0.0003	0.0018	1.0607	1.0553	0.9086
PgB17	354879	361743	CYP_4V2	0.0039	0.0029	0.0029	0.9553	-0.2165	-1.7553
PgR002	512210	524078	CYP_13A8	0.0002	0.0014	0.0051	-0.1263	1.1847	0.6614
PgR006	2812763	2821080	CYP_4V2	0.001	0.0005	0.0008	-0.3596	1.0642	0.8911
PgR010	2986365	2996675	CYP_3A31	0.0072	0.0086	0.0065	1.3057	2.4035	2.0071
PgR020	456556	464491	CYP_2C8	0.0068	0.0081	0.0091	0.7683	1.0655	0.6074
PgR027	284432	291992	CYP_2C25	0.0036	0.0043	0.0035	0.3752	1.6727	-0.4168
PgR033	836482	846733	CYP_4C1	0.0015	0.0107	0.0097	-2.6423	2.4022	1.5867
PgR049	713990	725852	CYP_4V2	0.0001	0.0005	0.0008	-1.4324	0.347	0.639
PgR012	2339498	2343997	CYP_3A7	0.0063	0.0125	0.0094	-1.2345	1.6788	0.0835

[1]note: β tubulin, β tubulin; glc-1, glutamate-gated chloride channel alpha; pgp, multidrug resistance-associated protein family; mrp, multidrug resistance protein family; haf, half transporter family; cup-4, acetylcholine receptor-like protein; bre-4, beta-n-acetylgalactosaminyltransferase bre-4; unc, Acetylcholine receptor family; nrf-6, Nose resistant to fluoxetine protein 6; ced-7, ABC transporter ced-7; aex-3, MAP kinase-activating death domain protein; CYP, cytochrome P450 family.

[2]PEc, *Parascaris* spp. worms from *Equus caballus* .

[3]PEz, *Parascaris* spp. worms from *Equus zebra* .

[4]PEa, *Parascaris* spp. worms from *Equus asinus* .

FIGURE LEGENDS

Fig.1 Sampling localities and geographical distribution of the three distinct populations of *Parascaris* spp. The upper left corner was a zebra-derived roundworm.

Fig.2 Population structure and relationships of roundworms from horse, zebra and donkey. (a) Principal component analysis (PCA) plots of the first three components. The fraction of the variance explained was 12.75% for PC1, 7.09% for PC2 and 0.06% for PC3; (b) Phylogenetic tree [maximum-likelihood (ML) tree with 1000 bootstraps] of all samples inferred from whole-genome tag SNPs, with *B. schroederi*as an outgroup; (c) Population structure plots with K=2-4. The y axis quantifies the proportion of the individual's genome from inferred ancestral populations, and x axis shows the different populations.

Fig.3 Chronogram of the *Parascaris* spp. based on Bayesian coalescent analysis of SNP data

using SNAPP. Nodes with high support (posterior probability = 1.00) are filled in red color. Error bars represent the 95% highest posterior densities (HPD). The colored circles represent different populations.

Fig.4 Demographic history of the *Parascaris* spp. reconstructed from the reference and population resequencing genomes.(a) The colored lines represent the estimated effective population size of each population. The 100 curves of each color represent the PSMC estimates for 100 sequences randomly resampled from the original sequence. The generation time (g) and the neutral mutation rate per generation (μ) of *Parascaris* spp. were 0.17 years and 0.9×10^{-8}, respectively. (b) Coalescent-based inference of demographic history using MSMC2. The upper panel shows the effective population sizes (Ne) of three populations, while the lower panel shows the split time between three populations; (c) Effective population size and split time based on SMC++ method.

Fig.5 Demographic inferences and early gene flow of*Parascaris* spp. populations. (a) Results of the population genetic model comparison using the three-dimensional site frequency spectrum (3D-SFS) between the PEc, PEz and PEa populations. A simplified graph of the best-fit model is depicted, along with the comparison of the 3D-SFS for data, model and residuals. (b) Results of the population genetic model comparison using the two-dimensional site frequency spectrum (2D-SFS) between PEc and PEz & PEa population along with the 2D-SFS for data, model and residuals.

Fig. 6 Schematic diagram of glycolysis, tricarboxylic acid cycle and lipid metabolism. The red arrow represents significantly positively selected enzymes in the PEz&PEa clade (P<0.01), and the blue arrow represents significantly positively selected enzymes in the PEc clade (P <0.01). The Manhattan plot is the XP-EHH score of the 50k region around the related genes.