Equus roundworms (*Parascaris univalens*) are undergoing rapid divergence due to host and environment factors

Lei Han

Northeast Forestry University

Tianming Lan

BGI-Shenzhen

Yaxian Lu

Northeast Forestry University

Mengchao Zhou

Northeast Forestry University

Haimeng Li

BGI-Shenzhen

Haorong Lu

BGI-Shenzhen

Qing Wang

BGI-Shenzhen

Xiuyun Li

Harbin northern forest zoo

Shan Du

Inner Mongolia Agriculture University Hohhot

Chunyu Guan

Harbin northern forest zoo

Yong Zhang

Center for Animal Disease Control and Prevention of Ordos

Sunil Kumar Sahu

BGI-Shenzhen

Puyi Qian

BGI-Shenzhen

Shaofang Zhang

BGI-Shenzhen

Hongcheng Zhou

BGI-Shenzhen

Wei Guo

Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences

Hongliang Chai

Northeast Forestry University

Sibo Wang

BGI-Shenzhen

Quan Liu

Northeast Forestry University

Huan Liu

BGI-Shenzhen

Zhijun Hou (✉ houzhijundb@163.com)

Northeast Forestry University

Research Article

Abstract

Background

The evolution of parasites is often directly affected by the host's environment. Studies on the evolution of the same parasites in different hosts are extremely attractive and highly relevant to our understanding of divergence and speciation.

Methods

Here we performed whole genome sequencing of *Parascaris univalens* from different Equus hosts (horses, zebras and donkeys). Phylogenetic and selection analysis was performed to study the divergence and adaptability of *P. univalens*.

Results

At the genetic level, multiple lines of evidence support that *P. univalens* were mainly separated into two clades (Horse-derived and Zebra & Donkey-derived). This divergence began at 300-1000 years ago, and we found that most of the key enzymes related to glycolysis were under strong positive selection in zebra & donkey-derived roundworms, but lipid related metabolism system was under positive selection in the horse-derived roundworms, indicating that the adaptive evolution of metabolism may drive the divergence in past few centuries. In addition, we found that some drug-related genes have a significantly higher degree of selection in different populations.

Conclusions

This work reports evidence that the host's diet drives the divergence of roundworms for the first time, and also supports that divergence is a continuous and dynamic process, and continuous monitoring of the effects of differences in nutritional and drug history on rapid evolution of roundworms are conducive to further understanding host-parasite interactions.

Introduction

Parascaris univalens 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 *P. univalens*. Despite their close relationship, the extent of habitats, food composition, digestion levels, and history of human intervention in these hosts are vastly different, and the impact on large parasites living in the host remains unknown. Comparatively, donkeys have much lower energy and protein requirements than those of other equids [1], and even the metabolic response during exercise will differ depending on the level of food given to the horse [2], 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 [3]. Roundworms' underlying genetic diversity contributes to their evolutionary adaptation. Understanding these processes, as well as the key environmental factors that drive the adaptation, is crucial to comprehend the specific evolutionary trends of parasites [4]. The genetic architecture and selection are key indicators of evolution aiming to understand how natural selection can lead to lineage divergence and speciation [5]. The genetic differences between populations can reveal a lot about the basic evolutionary process due to changes in the environment [6]. Studying the adaptation of parasites along with different hosts are of great significance for understanding the host-parasite interactions.

The control of parasitic nematodes feeding on animal relies almost exclusively on anthelmintic, 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 [7]. There is widespread concern about the risk associated with relying on anthelmintics with hundreds of millions of doses being donated and used every year [8]. 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 [9–11]. Due to the inconsistency of drug history in different regions or farms, roundworms are facing varying degrees of selection pressure. Timely monitoring of drug use on the evolution of roundworms is of great significance. In recent years, genome scanning has become an effective means to reveal the genetic determinants of adaptive evolution in different habitats 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 [12, 13]. Here we analyzed the genome characteristics of *P. univalens* populations dwelling on three main hosts: horse (*E. caballus*), zebra (*E. zebra*) and donkey (*E. asinus*) in northeastern China. We presented the first report on the recent divergence of *P. univalens* populations, and speculated that it might be linked to host digestion and metabolism preference. In addition, the role of selection pressure in the population was also evaluated at the genetic level. This work is of significance for changing the current "one size fits all" roundworm taxonomy, and providing a reference for revealing potential differentiation trends.

Results

Genome resequencing and genetic variations

A total of 42 individuals from three *P. univalens* of Equus (PEc, n=19), zebra (PEz, n=18), donkey (PEa, n=5) from Inner Mongolia and Heilongjiang of China were whole-genome sequenced (Fig. 1, Table S1). 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 was 98.17% and genomic coverages were larger than 90% for all individuals (Table S2). Currently, there may be two types of Equus roundworms due to different chromosome numbers (*P. equorum* n=4, or *P. univalens* n=2). We first performed species identification. We randomly selected 1-4 samples at each sampling site for karyotype identification and found that all samples had two chromosomes (see methods; Fig. S3). In addition, the BWA alignment confirmed that the all samples were closer to *P. univalens*, with an average mapping rate against *P. equorum* of 78%, significantly lower than that against the *P. univalens* genome (98.03%, P<0.01) (Fig. S4a). These results indicate that the collected samples were all *P. univalens*. In addition, we calculated the observed heterozygosity and the expected heterozygosity, indicating that the inbreeding coefficient between individuals were very low (Average F < 0.1), and the possible genetic relationships were relatively distant, which were representative (Table S3).

Genetic differentiation was found in the *P. univalens*

Principal components analysis (PCA) supported the clear separation among *P. univalens* populations (Fig. 2a), with PC1 and PC2 separating the PEc and PEz&PEa populations (P<0.05). All PCs showed that PEz and PEa were in the same cluster. In addition, 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 admixture 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 identity-by-descent (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 IBD fragments (Fig. S5, Fig. S6). In addition, lower inter-population *Fst* values were detected in PEz and PEa populations also indicated a closer relationship between them (Fig. S6).

Inference of demographic history and divergence time in *P. univalens* populations

To examine the genome-wide divergence time among *P. univalens* 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). Further Ne estimation inferred by the MSMC2 also supported the PSMC result with almost a similar trend 1000 years before, without differentiations, indicating the existence of possibly common ancestor (Fig. 4a, 4b). However, the Ne of the three populations began to diverge in recent ~800 years (Fig. 4b). Further, we found that the PEc population was genetically separated from the PEz and PEa (Fig. 4b) at ~ 300 years ago with the RCCR less than 0.5. However, there was no obvious sign of separation between PEz and PEa (RCCR >0.5). Separations among these three populations were further supported and validated by results inferred by SMC++, which was independent of phased genotypes (Fig. 4c). We carefully compared the relationship between the separation and the topology of the phylogenetic tree, and found similar results. Both results showed that PEc vs PEz, PEc vs PEa had obvious divergence, but PEz vs PEa are not fully differentiated. Although the divergence times estimated by the above methods are not completely consistent, they all indicate that this differentiation occurred recently. Taken together, we considered that *P. univalens* were mainly divided into two clades, one was from horse-derived (PEc) and the other one was from zebra & donkey-derived (PEz&PEa). The conservative divergence time was estimated within the recent 300 years. Finally, we used

fastsimcoal2 to evaluate the population size after divergence based on the observed joint site frequency spectrum (SFS; Fig. S7), and found that the PEc population size was significantly larger than PEz&PEa after this divergence event (Fig. 4d). The best coalescent simulation model inferred by fastsimcoal2 also indicates an early bidirectional gene flow between PEc and PEz&PEa.

The most possible demographic model in P. univalens **populations**

In order to better understand this recent divergence of *P. univalens*, we used the δaδi to further explore the demographic history of the divergence. Following the method of Portik [14], 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. S8). 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. S9), 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.

Different Adaptive Strategies From The Prospective Of Metabolism

In-depth genome scanning and functional annotation helped us to understand the population divergence. Genome-wide nucleotide diversity (π) was computed for each population with all samples. Meanwhile, we identified genomic regions as candidate divergent regions (CDRs) among PEc, PEz and PEa populations (Table S6, Fig. S10). 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. S11). These 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. S12). In addition, we also used the XP-EHH method to screen genes that may have been positively selected in 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. S13, Fig. S14, Table S8). PEz&PEa clade has shown significant positive selection in almost all key enzymes in glycolysis and tricarboxylic acid cycle. These loci have aroused our attention and we 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 [15], largely due to their anaerobic mode of life. The significant selection signals in 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.

Anthelmintics, The Potential Drivers Of Population 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 [16] 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 main reported so far (Table 1). We also scanned iHS scores 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 under strong positive selection among different populations (Table 1). The genetic diversity of some drug-related gene regions was significantly low and conservative. We identified three classic resistance locus (167/Phe, 198/Glu and 200/Phe) of the BZs resistance gene β-tubulin [17], and found that all individuals in the three populations did not show resistant mutations through sequence alignment (Fig. S15a). However, these gene regions were found under strong recently positive selection (Fig. S15c). In addition, the selection signals in other resistance-related genes in the population were also discovered (Fig. S16 - S18). For example, multidrug resistance protein pgp-3 and multidrug resistance protein 1 (*mrp-1*) were found to be under 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. A fluoxetine (Prozac) resistance gene nose resistant to fluoxetine protein 6 (*nrf-6*) [18] showed strong positive selection in three populations. Positive selections of these gene regions were different in three populations, which may be related to the drug history in different environments. Nevertheless, it was obvious that in some populations, certain genetic selections were very conservative and significant. It can be seen from the results that the selection of these resistance-related genes will be retained in independent populations and would continue to be passed down to the next generation. It should be noted that these selections were population-specific and have the potential to promote population differentiation.

Table 1. Nucleotide diversity (π) and nature selections around the key resistance-related proteins' region in PEc, PEz and PEa populations.

Scaffold ID	Start	End	Description[1]	π			Tajima'D			Significantly selected populations
				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	PEc, PEz, PEa
PgR004	1320231	1320527	*glc-1*	0.0021	0.0017	0.0007	-0.4035	2.0857	-1.8797	PEz, PEa
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	PEc
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	PEc
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	PEc
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	PEc, PEz, PEa
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	PEz, PEa
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	PEz
PgR033	836482	846733	CYP_4C1	0.0015	0.0107	0.0097	-2.6423	2.4022	1.5867	PEc
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, *P. univalens* from *Equus caballus*.

[3]PEz, *P. univalens* from *Equus zebra*.

[4]PEa, *P. univalens* from *Equus asinus*.

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 evolution of a single species is a long and delicate process. Numerous environmental changes and its own factors affect this process. 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 simulate the most possible demographic history. 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 *P. univalens* populations. In addition, the demographic history also showed that the *P. univalens* were in the process of divergence. Recent selection analysis provided evidence for understanding the possible causes of the divergence, which supported the significant impact of different host intestinal habitats on evolution. We summarized the main findings on the diversification of *P. univalens*, and provided a perspective for future monitoring of roundworm ecology in a timely manner to deal with possible unfavorable mutations.

For parasites, the host's influence on its evolution cannot be ignored. 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 PEz and PEa were subjected to a higher degree of recent positive 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 diet. 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 [19]. Immunological evasion could be the major purpose. This notion was 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 [20]. 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 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 [21], 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 [22]. In our results, we found that some resistance-related genes such as β-tubulin, glc-1, pgp-3, mrp-6, cup-4, nrf-6 and CYP family, were under strong positive selection 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. Currently, in addition to anthelmintics, vaccines also offer an attractive alternate control strategy against these parasites [23]. 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. Compared with non-human interference natural selection, the effect of anthelmintics is undoubtedly more direct and stronger, and the consequences of this choice are difficult to predict. This suggests that we should be cautious in dealing with the issue of drug resistance and adopt more scientific strategies. It should be noted that *P. univalens* are highly adaptable and evolve quickly. As the gene frequency of certain drug-related genes in specific populations further increases, it was likely that the *P. univalens* will differentiate into new subpopulations with specific phenotypes. Long-term use of anthelmintics will become a potential driver for the differentiation of *P. univalens*.

This work provides a reference for monitoring the evolution of parasites and revealing the driving factors of evolution under the action of natural selection and drugs. It should be noted that due to the limitation of sample size, some accidental or other environmental factors cannot be fully considered for natural selection in *P. univalens*. We only considered two types of extremely significant factors (host's diet and anthelmintics), and other selection effects still need to be verified in a wider range of *P. univalens* in

the future. It is foreseeable that the impact of human intervention in the host-parasite interaction will become greater and greater, and how to find a balance point will still be a major challenge in the future.

Methods

Sample collection

Seventeen roundworms were collected from six naturally infected horses (*E. 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 roundworms from three donkey (*E. asinus*) were collected after anthelmintic treatment from a farm in Chifeng, Inner Mongolia, China. Eighteen roundworms from ten zebra (*E. zebra*) were obtained after anthelmintics treated from Harbin Northern Forest Zoo, Heilongjiang, China. All specimens were washed extensively in sterile physiological saline (37 °C), snap-frozen and transported with dry ice and then stored at -80 °C until further use.

Karyotyping

We performed karyotyping on the collected samples. Worms were carefully dissected and the gonads located and excised. The gonads were then processed for karyotyping as previously described [24]. Using a modified freeze-crack method to permeabilize and fix embryos. Briefly, the embryos were immersed in KCl (0.075 M) hypotonic 5 min, then rinse with methanol/acetic (3:1) acid solution. Next, drip 45% acetic acid on the siliconized coverslip. After pressurizing for about 60 seconds, put the slider in liquid nitrogen and freeze for 1-2 minutes. Staining was carried out with 4',6-diamidino-2-phenylindole (DAPI) for 5 min, and the slides were then examined under a fluorescence microscope.

Nucleic acid isolation, library construction and sequencing

Total genomic DNA was isolated using a sodium dodecyl sulphate/proteinase K digestion [25] 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).

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) [26] with default parameters. SAMtools (v0.1.19) [27] 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) [28] 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) [29] for downstream bioinformatic analyses. Variants Annotation of variants were 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.

Diversity analysis

SNP density was counted with a 10 kb sliding window using VCFtools (v0.1.13) software [29]. 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).

Phylogenetic relationship, genetic structure and admixture

Principal component analysis (PCA) was carried out using EIGENSOFT [30] software base on the SNP dataset, and the population clustering analysis was conducted in PLINK (v1.90b6.10) [31]. We used genome-wide SNPs to construct the maximum likelihood (ML) phylogenetic tree with 1000 bootstraps using iqtree (v1.6.12) [32]. 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 *P. univalens* populations in a coalescent framework with SNAPP implemented in BEAST v2.6.3 [33]. 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 [34]. According to epidemiological investigation, we assumed that the average generation time of *P. univalen* sroundworms 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.

Estimates of the effective population size and divergence time

The pairwise sequentially Markovian coalescent (PSMC) method [35] 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 [36]. 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 multiple sequentially Markovian coalescent (MSMC2) [37], 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 MSMC2, we also used Sequentially Markovian Coalescent (SMC++) methods [38] 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) [39], then the calculation was performed with the following parameters: -i 20 -t 6 -p '10*1+15*2'. The mutation rate (μ) of *P. univalens* for SMC++ and MSMC2 were used the same values as for PSMC.

Demographic inference using fastsimcoal2 and δaδi

We used the fastsimcoal2 [40] approach to deduce the recent demographic history of *P. univalens*populations. We chose only SNPs located in intergenic regions to avoid the influence of SNPs under selection [41]. We used δaδi [42] 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.

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 [43]. We use SNPs with an iHS score of top 0.5% and the distance between adjacent SNPs < 50 kb as candidate SNPs [44]. 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 cross population extended haplotype homozygosity (XP-EHH) method on each pair of combination (PEc vs PEz, PEc vs 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.

Conclusions

P. univalens is the main parasitic pathogen that infects equine, and it is also the chief culprit in horses' weight loss and weakened immunity. The genetic variation and host differences complicate the development of broad-spectrum diagnostics, therapeutic. Here we report the recent divergence of *P. univalens* and reveal that the rapid evolution of glycolysis-related genes drove this divergence. It is also a key factor leading to the parasitic preference of roundworm populations. In addition, we found that resistance-related genes have similar tendency, which was the potential impact of overuse of anthelmintics. We have established a rapid evolution gene set of

P. univalens, which will help managers decide on therapeutic strategies targeting specific populations, and allow researchers to monitor the ongoing evolution and diversification of *P. univalens*.

Declarations

Ethics approval and consent to participate

All experimental designs and nematode handling were approved by the Institutional Animal Care and Use Committee of Northeast Forestry University and in accordance with the laboratory of Entomopathogenic Diseases and Pathogen Ecology.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) [45] of China National GeneBank DataBase (CNGBdb) [46] with accession number CNP0001875. All in-house scripts and codes used in this study were available in the github database (https://github.com/HanLei12321/Equus_roundworms).

Competing interests

The authors declare no conflict financial interests.

Author contribution

Zhijun Hou, Huan Liu and Tianming Lan designed and initiated the project. Yaxian Lu, Xiuyun Li, Shan Du Chunyu Guan, Yong Zhang, and Lei Han collected the samples. Lei Han, Shaofang Zhang, Puyi Qian, Hongcheng Zhou, and Haorong Lu performed the DNA extraction, library construction and sequencing. Lei Han, Tianming Lan, Haimeng Li, and Qing Wang performed the data analysis. Lei Han, and Tianming Lan wrote the manuscript. Zhijun Hou, Huan Liu, Tianming Lan, Quan Liu, Sibo Wang, Wei Guo, Sunil Kumar Sahu, and Yaxian Lu supervised the manuscript. All authors read and approved the final manuscript.

References

1. Martin-Rosset W: **Donkey nutrition and feeding: Nutrient requirements and recommended allowances—A review and prospect**. *Journal of equine veterinary science* 2018, **65**:75–85.
2. Jansson A, Lindberg JE: **A forage-only diet alters the metabolic response of horses in training**. *Animal* 2012, **6**(12):1939–1946.
3. Hay SI, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abdulkader RS, Abdulle AM, Abebo TA, Abera SF: **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* 2017, **390**(10100):1260–1344.

4. Salle G, Doyle SR, Cortet J, Cabaret J, Berriman M, Holroyd N, Cotton JA: **The global diversity of Haemonchus contortus is shaped by human intervention and climate**. *Nature communications* 2019, **10**(1):1–14.

5. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S: **The genomic basis of adaptive evolution in threespine sticklebacks**. *Nature* 2012, **484**(7392):55–61.

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

7. Kaplan RM, Vidyashankar AN: **An inconvenient truth: global worming and anthelmintic resistance**. *Veterinary parasitology* 2012, **186**(1-2):70–78.

8. McManus D: **Schistosomiasis. Nature reviews. Disease primers 4, 13**. In.; 2018.

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

10. Kaplan RM, Klei TR, Lyons ET, Lester G, Courtney CH, French DD, Tolliver SC, Vidyashankar AN, Zhao Y: **Prevalence of anthelmintic resistant cyathostomes on horse farms**. *Journal of the American Veterinary Medical Association* 2004, **225**(6):903–910.

11. Traversa D, von Samson-Himmelstjerna G, Demeler J, Milillo P, Schürmann S, Barnes H, Otranto D, Perrucci S, di Regalbono AF, Beraldo P: **Anthelmintic resistance in cyathostomin populations from horse yards in Italy, United Kingdom and Germany**. In: *Parasites & Vectors: 2009*: Springer; 2009: 1–7.

12. Wu CI: **The genic view of the process of speciation**. *Journal of evolutionary biology* 2001, **14**(6):851–865.

13. Nosil P, Vines TH, Funk DJ: **Reproductive isolation caused by natural selection against immigrants from divergent habitats**. *Evolution* 2005, **59**(4):705–719.

14. Portik DM, Leaché AD, Rivera D, Barej MF, Burger M, Hirschfeld M, Rödel MO, Blackburn DC, Fujita MK: **Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection**. *Molecular ecology* 2017, **26**(19):5245–5263.

15. Frayha G, Smyth J: **Lipid metabolism in parasitic helminths**. *Advances in Parasitology* 1983, **22**:309–387.

16. Wolf PG, Rowe CA, Kinosian SP, Der JP, Lockhart PJ, Shepherd LD, McLenachan PA, Thomson JA: **Worldwide relationships in the fern genus Pteridium (bracken) based on nuclear genome markers**. *Am J Bot* 2019, **106**(10):1365–1376.

17. Lake SL, Matthews JB, Kaplan RM, Hodgkinson JE: **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**. In: *Parasites & vectors: 2009*: BioMed Central; 2009: 1–12.

18. Choy RKM, Kemner JM, Thomas JH: **Fluoxetine-Resistance Genes in Caenorhabditis elegans Function in the Intestine and May Act in Drug Transport**. *Genetics* 2006, **172**(2):885–892.

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

20. Wallis RC: **Current Topics in Insect Endocrinology and Nutrition**. *The Yale journal of biology and medicine* 1982, **54**(6):503–504.

21. Wang J, Davis RE: **Ascaris**. *Current biology: CB* 2020, **30**(10):R423-R425.

22. McKellar QA, Jackson F: **Veterinary anthelmintics: old and new**. *TRENDS in Parasitology* 2004, **20**(10):456–461.

23. Hewitson JP, Maizels RM: **Vaccination against helminth parasite infections**. *Expert review of vaccines* 2014, **13**(4):473–487.

24. Nielsen MK, Wang J, Davis R, Bellaw JL, Lyons ET, Lear TL, Goday C: **Parascaris univalens–a victim of large-scale misidentification?** *Parasitology Research* 2014, **113**(12):4485–4490.

25. Gasser RB, Hu M, Chilton NB, Campbell BE, Jex AJ, Otranto D, Cafarchia C, Beveridge I, Zhu X: **Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation**. *Nature Protocols* 2006, **1**(6):3121–3128.

26. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. 2010.

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

28. Depristo MA, Banks E, Poplin R, Garimella KV, Daly MJ: **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. *Nature Genetics* 2011, **43**(5):491–498.

29. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST: **The variant call format and VCFtools**. *Bioinformatics* 2011, **27**(15):2156–2158.

30. Nick, Patterson, Alkes L, Price, David, Reich: **Population structure and eigenanalysis**. *Plos Genetics* 2006.

31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIWD, Daly MJ: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *American Journal of Human Genetics* 2007, **81**(3):559–575.

32. Lam-Tung N, Schmidt HA, Arndt VH, Quang MB: **IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies**. *Molecular Biology & Evolution* 2015(1):268–274.

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

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

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

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

37. Schiffels S, Durbin R: **Inferring human population size and separation history from multiple genome sequences**. *Nature genetics* 2014, **46**(8):919–925.

38. Terhorst J, Kamm JA, Song YS: **Robust and scalable inference of population history from hundreds of unphased whole genomes**. *Nature genetics* 2017, **49**(2):303–309.

39. Browning SR, Browning BL: **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* 2007, **81**(5):1084–1097.

40. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M: **Robust demographic inference from genomic and SNP data**. *PLoS Genet* 2013, **9**(10):e1003905.

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

42. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD: **Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data**. *PLoS genet* 2009, **5**(10):e1000695.

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

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

45. Guo X, Chen F, Gao F, Li L, Liu K, You L, Hua C, Yang F, Liu W, Peng C: **CNSA: a data repository for archiving omics data**. *Database* 2020, **2020**(2020):baaa055.

46. Feng ZC, Li JY, Fan Y, Li NW, Xiao FW: **CNGBdb: China National GeneBank DataBase**. *Hereditas* 2020, **42**(8):799–809.
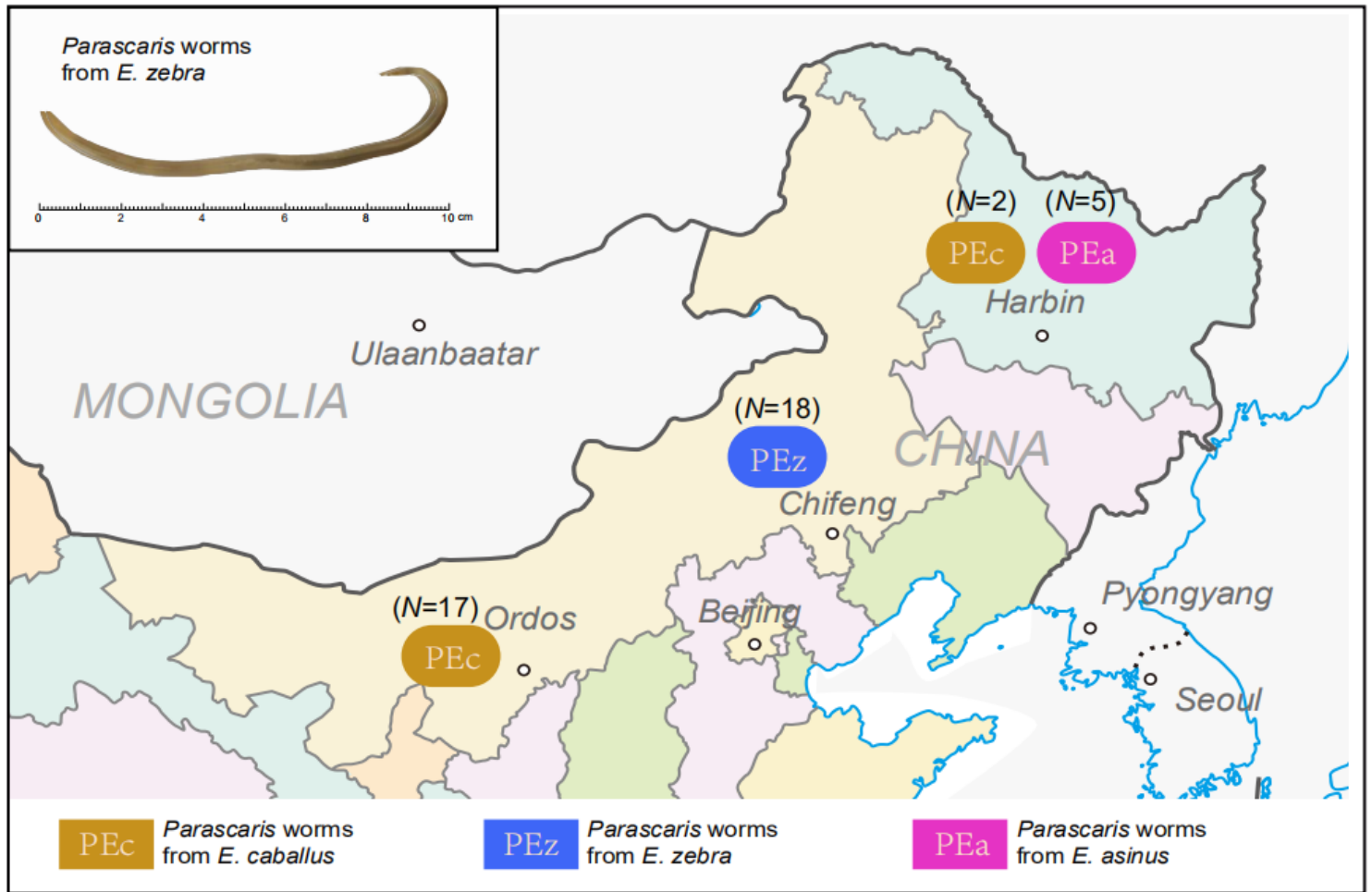
Figures

Figure 1

Sampling localities and geographical distribution of the three distinct populations of P. univalens. The upper left corner was a zebra-derived roundworm.
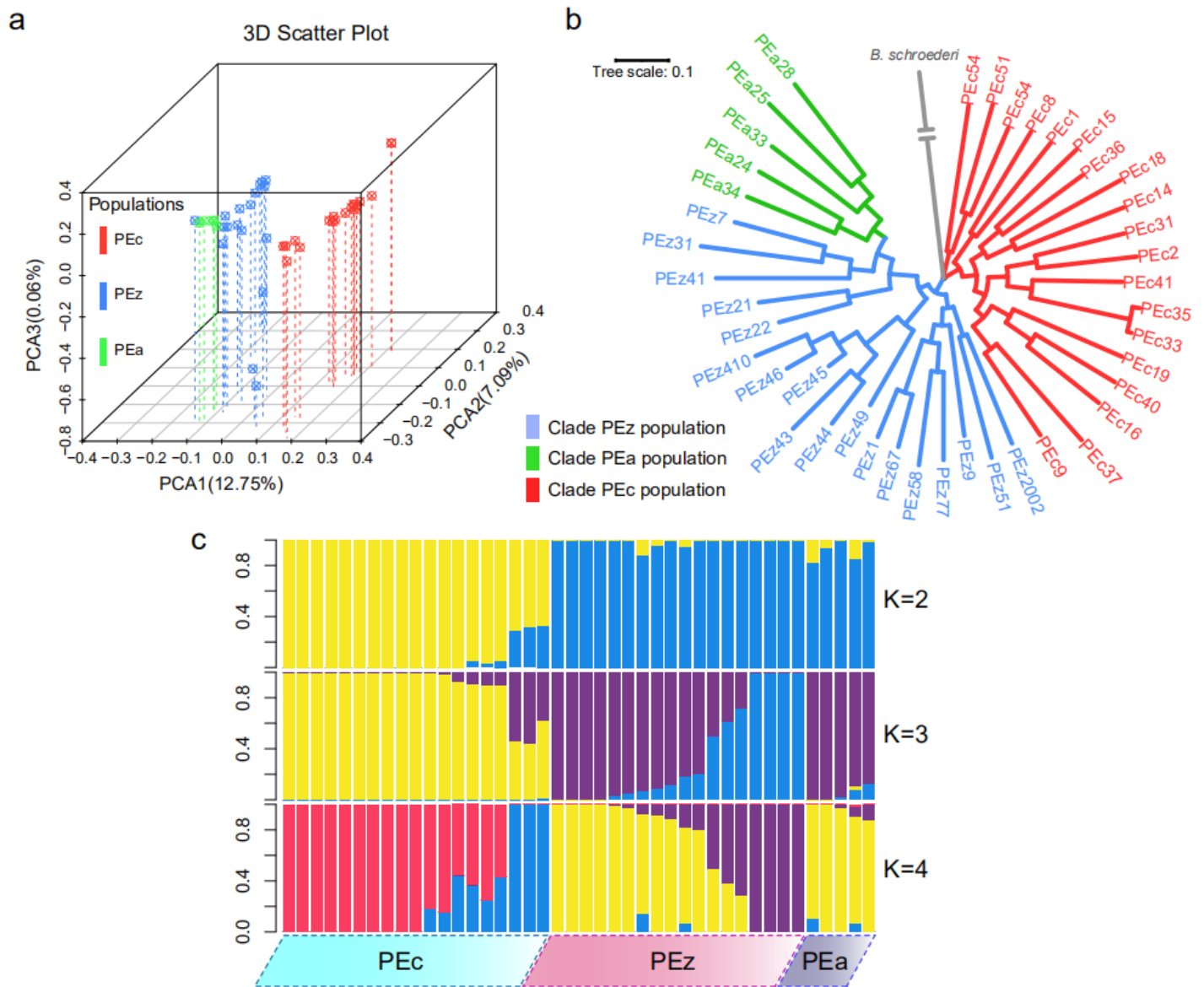
Figure 2

Population structure and relationships of P. univalens from horse, zebra and donkey. (a) 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 (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.
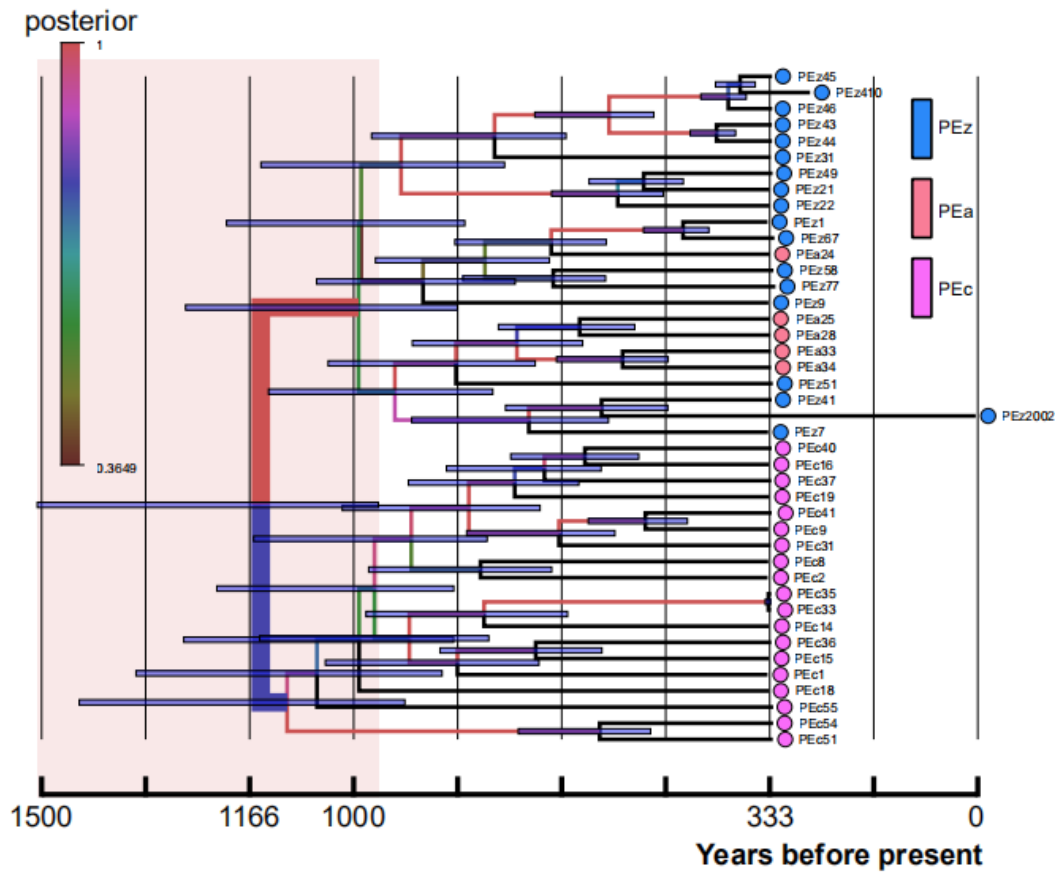
Figure 3

Chronogram of the P. univalens 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.
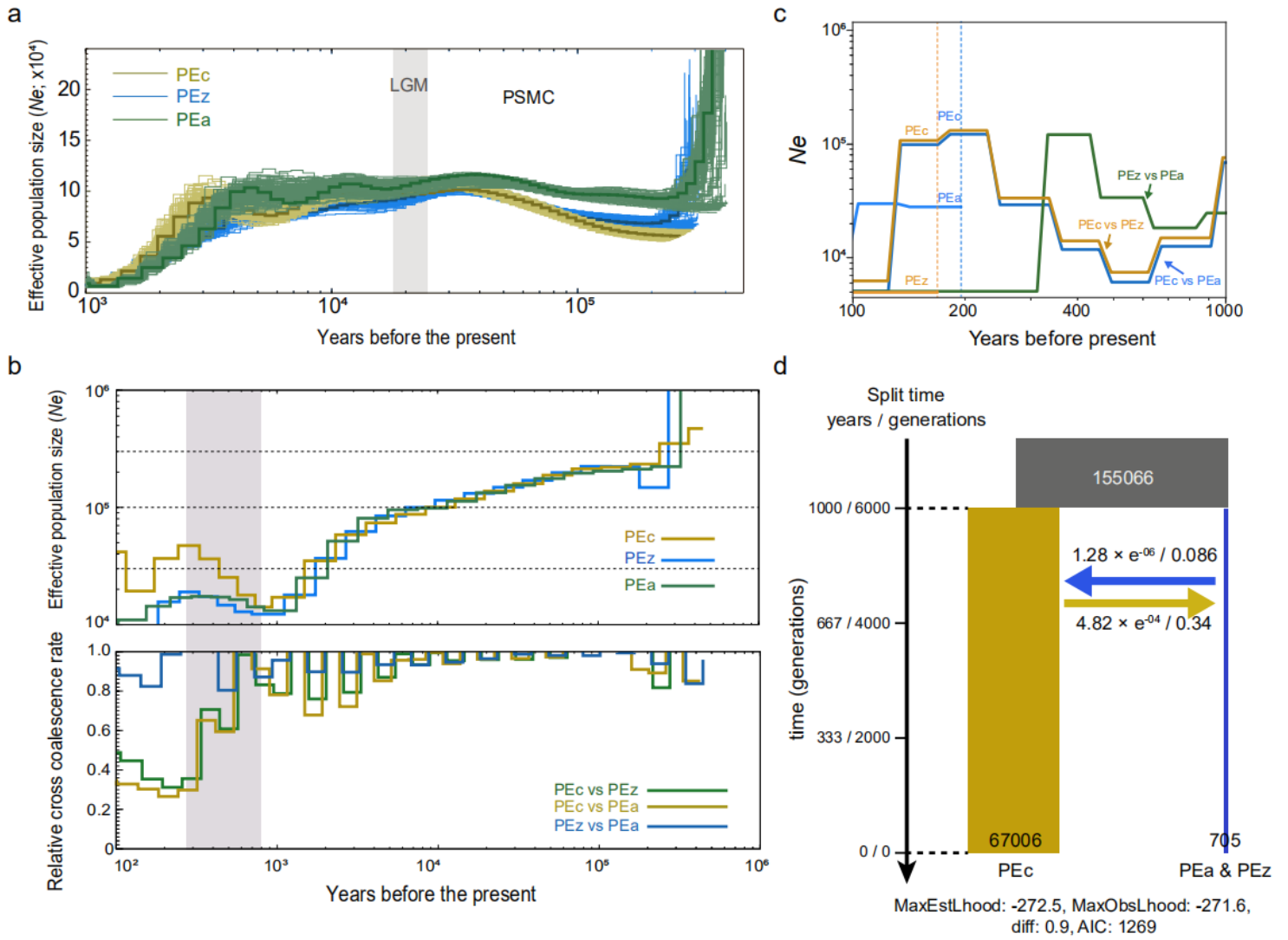
Figure 4

Demographic history of the P. univalens populations 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 P. univalens 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.
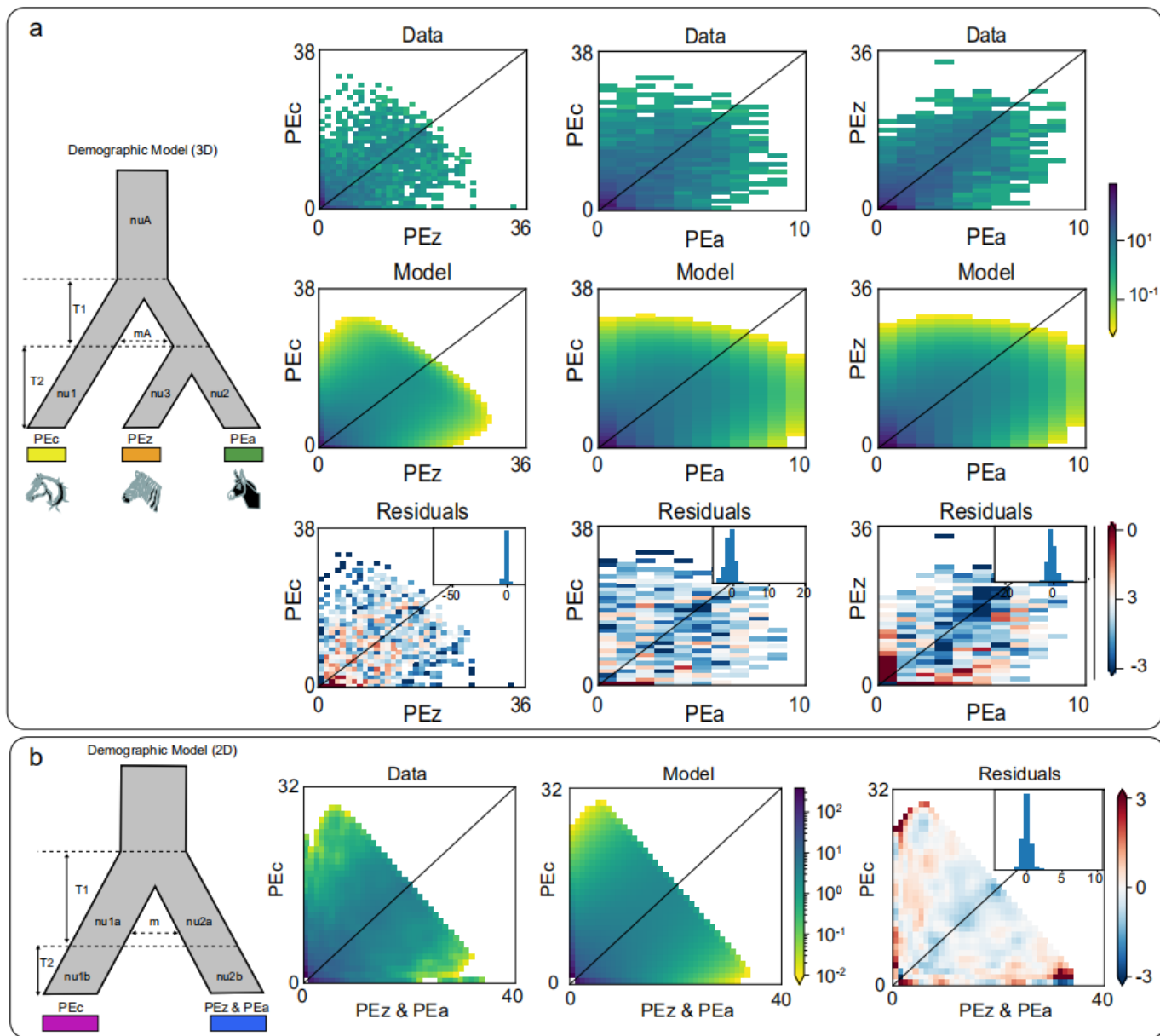
Figure 5

Demographic inferences and early gene flow of P. univalens 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, models and residuals.
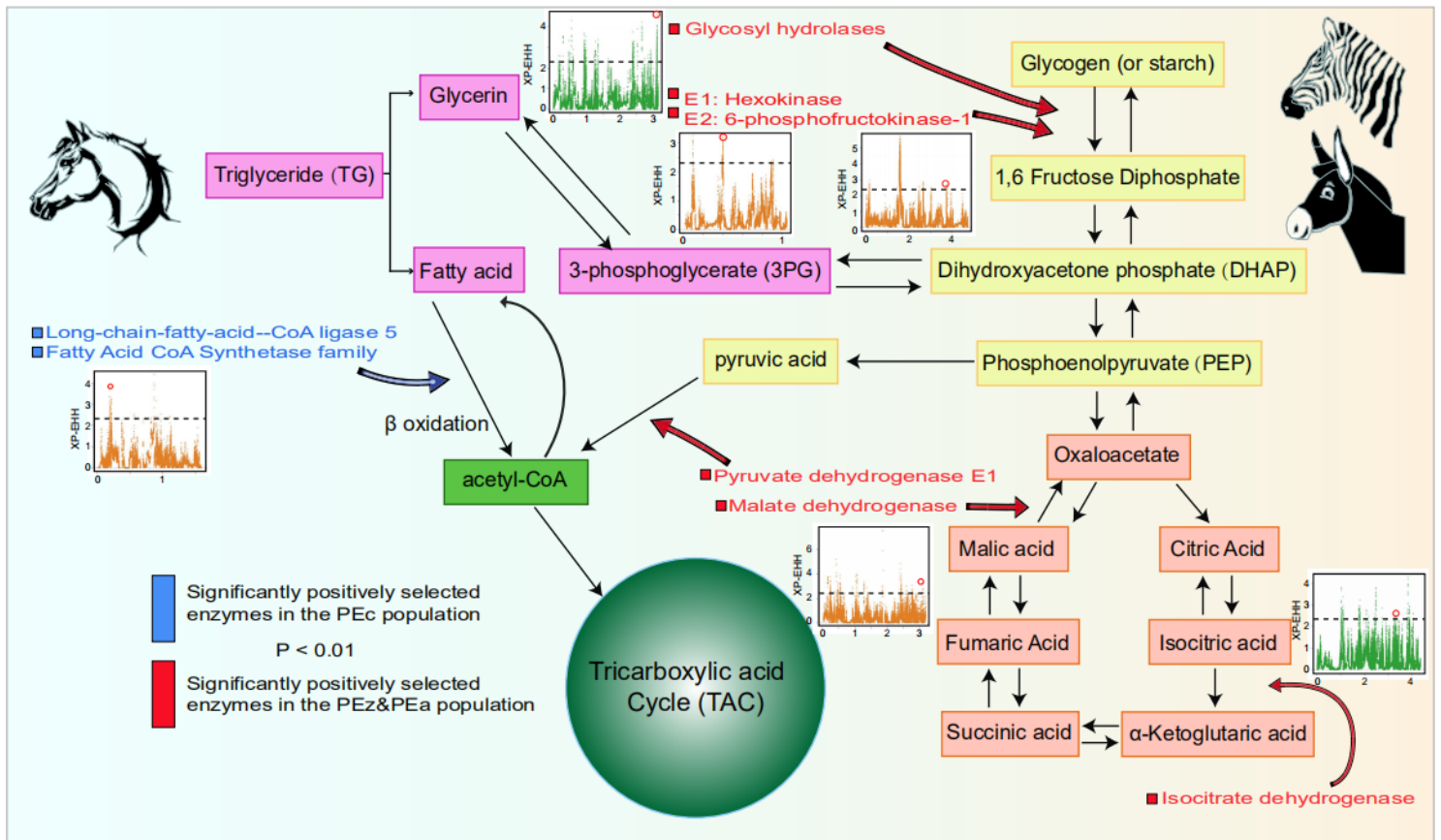
Figure 6

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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx