

# Chromosome-scale assembly and whole-genome sequencing of 266 giant panda roundworms provide insights into their evolution, adaptation and potential drug targets

Lei Han<sup>1</sup>, Tianming Lan<sup>2</sup>, Desheng Li<sup>3</sup>, Haimeng Li<sup>4</sup>, Linhua Deng<sup>5</sup>, Zhiwei Peng<sup>3</sup>, Shaowen He<sup>6</sup>, Yanqiang Zhou<sup>1</sup>, Ruobing Han<sup>1</sup>, Lingling Li<sup>1</sup>, Yaxian Lu<sup>1</sup>, Haorong Lu<sup>2</sup>, Qing Wang<sup>7</sup>, Shangchen Yang<sup>8</sup>, Yixin Zhu<sup>2</sup>, Yunting Huang<sup>2</sup>, Xiaofang Cheng<sup>9</sup>, Jieyao Yu<sup>2</sup>, Yulong Wang<sup>1</sup>, Heting Sun<sup>1</sup>, Hongliang Chai<sup>1</sup>, Huanming Yang<sup>2</sup>, Xun Xu<sup>10</sup>, Michael Lisby<sup>11</sup>, Quan Liu<sup>1</sup>, Karsten Kristiansen<sup>3</sup>, Huan Liu<sup>12</sup>, and Zhijun Hou<sup>1</sup>

<sup>1</sup>Northeast Forestry University

<sup>2</sup>State Key Laboratory of Agricultural Genomics, BGI-Shenzhen

<sup>3</sup>Affiliation not available

<sup>4</sup>BGI-Shenzhen

<sup>5</sup>Key Laboratory of SFGA on Conservation Biology of Rare Animals in the Giant Panda National Park (CCRCGP)

<sup>6</sup>Foping National Nature Reserve

<sup>7</sup>University of Chinese Academy of Sciences

<sup>8</sup>Zhejiang University

<sup>9</sup>MGI, BGI-Shenzhen

<sup>10</sup>BGI

<sup>11</sup>Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen

<sup>12</sup>University of Copenhagen

April 9, 2021

## Abstract

Helminth diseases have long been a threat to the health of humans and animals. Roundworms are important organisms for studying parasitic mechanisms, disease transmission and prevention. The study of parasites in the living fossil giant panda is of great significance for understanding the adaptation mechanism of roundworms to the host. Here, we report a high-quality chromosome-scale genome of *Baylisascaris schroederi* with a genome size of 262 Mb and 19,291 predicted protein-coding genes. We found a significant expansion of genes related to epidermal chitin synthesis and environmental information processing in roundworms genome. Furthermore, we demonstrated unique genes involved in essential amino acid metabolism in the *B. schroederi* genome, inferred to be essential for the adaptation to the giant panda-specific diet. In addition, under different deworming pressures, we found that four resistance-related genes (*glc-1*, *nrf-6*, *bre-4* and *ced-7*) were under strong positive selection in captive population. Finally, 23 known drug targets and 47 potential target proteins were identified. The genome provides a unique reference for inferring the early evolution of roundworms and the mechanisms underlying adaptive. Population genetic analysis and drug prediction provide insights for revealing the impact of deworming history on population genetic structure and prevention.

Chromosome-scale assembly and whole-genome sequencing of 266 giant panda roundworms provide insights into their evolution, adaptation and potential drug targets

Lei Han<sup>1,5+</sup>, Tianming Lan<sup>2,3+</sup>, Desheng Li<sup>4+</sup>, Haimeng Li<sup>2,7</sup>, Linhua Deng<sup>4</sup>, Zhiwei Peng<sup>1</sup>, Shaowen He<sup>15</sup>, Yanqiang Zhou<sup>1</sup>, Ruobing Han<sup>1</sup>, Lingling Li<sup>1</sup>, Yaxian Lu<sup>1</sup>, Haorong Lu<sup>8,9</sup>, Qing Wang<sup>12</sup>, Shangchen Yang<sup>10</sup>, Yixin Zhu<sup>12</sup>, Yunting Huang<sup>8,9</sup>, Xiaofang Cheng<sup>11</sup>, Jieyao Yu<sup>8,9</sup>, Yulong Wang<sup>1</sup>, Heting Sun<sup>13</sup>, Hongliang Chai<sup>1</sup>, Huanming Yang<sup>2</sup>, Xun Xu<sup>2,8</sup>, Michael Lisby<sup>3</sup>, Quan Liu<sup>1,6\*</sup>, Karsten Kristiansen<sup>3,14\*</sup>, Huan Liu<sup>2,3\*</sup>, Zhijun Hou<sup>1,5\*</sup>

<sup>1</sup>College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China

<sup>2</sup>State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China

<sup>3</sup>Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

<sup>4</sup>Key Laboratory of SFGA on Conservation Biology of Rare Animals in the Giant Panda National Park (CCRCGP), Sichuan 611800, China

<sup>5</sup>Key Laboratory of Wildlife Conservation, China State Forestry Administration, Harbin 150040, China

<sup>6</sup>School of Life Sciences and Engineering, Foshan University, Foshan, Guangdong Province, China

<sup>7</sup>School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>8</sup>Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen 518120, China

<sup>9</sup>China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China.

<sup>10</sup>College of Life Sciences, Zhejiang University, Hangzhou 310058, China

<sup>11</sup>MGI, BGI-Shenzhen, Shenzhen 518083, China

<sup>12</sup>BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China

<sup>13</sup>General Station for Surveillance of Wildlife Diseases, National Forestry and Grassland Administration, Harbin 150040, China

<sup>14</sup>Qingdao-Europe Advanced Institute for Life Sciences, Qingdao 266555, China

<sup>15</sup>Foping National Nature Reserve, Hanzhong 723400, China

<sup>+</sup>These authors contributed equally to this work.

\*Corresponding authors: [houshijundz@163.com](mailto:houshijundz@163.com), [liuhuan@genomics.cn](mailto:liuhuan@genomics.cn), [kk@bio.ku.dk](mailto:kk@bio.ku.dk) and [liuquan1973@hotmail.com](mailto:liuquan1973@hotmail.com)

## Abstract

Helminth diseases have long been a threat to the health of humans and animals. Roundworms are important organisms for studying parasitic mechanisms, disease transmission and prevention. The study of parasites in the living fossil giant panda is of great significance for understanding the adaptation mechanism of roundworms to the host. Here, we report a high-quality chromosome-scale genome of *Baylisascaris schroederi* with a genome size of 262 Mb and 19,291 predicted protein-coding genes. We found a significant expansion of genes related to epidermal chitin synthesis and environmental information processing in roundworms genome. Furthermore, we demonstrated unique genes involved in essential amino acid metabolism in the *B. schroederi* genome, inferred to be essential for the adaptation to the giant panda-specific diet. In addition, under different deworming pressures, we found that four resistance-related genes (*glc-1*, *nrf-6*, *bre-4* and *ced-7*) were under strong positive selection in captive population. Finally, 23 known drug targets and 47 potential target proteins were identified. The genome provides a unique reference for inferring the early evolution of roundworms and the mechanisms underlying adaptive. Population genetic analysis and drug

prediction provide insights for revealing the impact of deworming history on population genetic structure and prevention.

**Keywords:** *Baylisascaris schroederi* , Roundworms, Adaptation, Genetic diversity, Anthelmintics

## 1 | INTRODUCTION

Parasitic ascariasis has long been a threat to the health of humans, livestock and wildlife worldwide (Hotez, Fenwick, Savioli, & Molyneux, 2009). With the expansion of towns, cities, and the wild land-urban interface, geographic isolation is no longer an effective barrier for transmission of helminth infections. As a result, the risks for transmission of diseases once isolated in wildlife have never been greater (K. Kazacos & W. M. Boyce, 1989). Due to its wide distribution and long incubation period, soil-transmitted helminth eggs are easily transmitted between wildlife and livestock, and even to humans through contaminated feces or soil. In-depth studies of helminths in wildlife can provide information of relevance for identifying and detecting pathogens and instigate appropriate actions to deal with possible risks with broad and far-reaching implications for wildlife and human health.

*B. schroederi* , a parasitic nematode specific for the giant panda (*Ailuropoda melanoleuca* ), is a soil-transmitted nematode and can directly infect the giant panda without passing through an intermediate host (Bethony et al., 2006; De Silva et al., 2003). *Baylisascaris* species also cause infection as patent or latent larval migrants (LM) in a variety of mammals (K. Kazacos & W. M. Boyce, 1989), birds (Wolf, Lock, Carpenter, & Garner, 2007) and humans (Murray, 2002; Wise, Sorvillo, Shafir, Ash, & Berlin, 2005), and are therefore considered zoonotic parasites with potential public health and safety risks. Its eggs can develop directly into infective larvae under appropriate conditions. Larvae can migrate to multiple organs of the host and cause visceral larval migrans (VLM), ocular larva migrans (OLM), neural larva migrans (NLM) and even severe pneumonia and hepatitis (K. R. Kazacos & W. M. Boyce, 1989; Papini, Renzoni, Malloggi, & Casarosa, 1995; Wildt, Zhang, Zhang, Janssen, & Ellis, 2006; L. Zhang et al., 2011). A large number of roundworm infections can in addition cause severe baylisascariasis, intestinal blockage, and even fatal bowel rupture (Schaul, 2006; GY Yang, 1998). Compared with other roundworms, *B. schroederi* is smaller in size and is mainly found in the small intestine of giant pandas. Giant pandas have typical carnivorous intestinal characteristics, but eat bamboo, a diet with low digestibility and absorption. This challenges the nutrient absorption of *B. schroederi* for survival in the small intestine. Based on available epidemiological data of the giant panda, *B. schroederi* is the leading cause of death from primary and secondary infection in wild and captive populations (H. Hu et al., 2018; D. Li et al., 2014). Moreover, the problem of increased resistance to anthelmintics is likely to be seriously underestimated. Giant pandas in captivity are regularly dewormed (every 60 days). According to investigations, *B. schroederi* eggs can still be detected in the feces 10 to 15 days after treatment with anthelmintics (D. Li, He, & Deng, 2015), indicating that a development of drug-resistant subtypes had occurred in the *B. schroederi* population, and that *B. schroederi* variants with resistance to a variety of anthelmintics had survived. These variants may potentially become anthelmintic-resistant pathogens.

Although *B. schroederi* poses threats to both wild and captive giant pandas, current studies are limited to morphological and single or multiple gene analyses, preventing in-depth exploration of genetic mechanisms of adaptations and further prevention and control of infections (Xie et al.). Here, we report a chromosome-scale reference genome of *B. schroederi* , which is also the first chromosome-scale reference genome of ascaridoids. Based on the genome, we explored possible genetic mechanisms of the adaptation of *B. schroederi* to the intestinal environment, especially the specific bamboo diet of the giant panda, as well as the potential genetic basis of drug resistance. Finally, potential drug target proteins were identified, which provides new insights into the potential disease management of *Baylisascaris* and related roundworms.

## 2 | MATERIALS AND METHODS

### 2.1 | Samples

All specimens of Sichuan *B. schroederi* population were collected from the giant panda (*Ailuropoda*

*melanoleuca* ) with naturally acquired infections from the China Conservation and Research Center for Giant Panda. Samples of Qinling *B. schroederi* population were obtained from the intestines of individuals who died after wild panda rescues in the Foping National Nature Reserve (Qinling Mountains) failed. Roundworms were washed extensively in sterile physiological saline (37 °C), sexes separated, snap-frozen and transport with dry ice and then stored at -80 °C until use. Several specimens were stored in RNA preservation solution for transcriptome sequencing. All experimental designs and nematodes handling were approved by the Institutional Animal Care and Use Committee of Northeast Forestry University.

## 2.2 | Nucleic acid isolation, library construction and sequencing

We firstly removed the intestines of adult individuals under an inverted microscope to prevent intestinal microorganism contamination. Then, total genomic DNA was isolated from them using a sodium-dodecyl sulphate/proteinase K digestion (Gasser et al., 2006) followed by phenol-chloroform extraction and ethanol precipitation. For PacBio sequencing, 20-Kb SMRTbell libraries were constructed using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, USA) and the SMRTbell Damage Repair Kit (Pacific Biosciences), according to the manufacturer’s instructions. The Pacbio sequencing was performed with 1 SMRT cell on the PacBio SMRTplatform. For genome resequencing, 100-bp pair-end libraries were constructed and sequenced a whole genome sequence (WGS) short reads on DNBSEQ-T1 platform. For RNA-seq, RNA was isolated separately from an adult female and a male individual of *B. schroederi* using TRIzol reagent (Invitrogen, USA) according to the manufacturer’s instructions and RNA yields were estimated spectrophotometrically (NanoDrop 1000), and integrity was verified using a Bioanalyzer. A total of 22Gb raw fastq data of female and 27 Gb raw fastq data of male were obtained. All transcripts were used to assess the completeness of the genome assembly and assist genome annotation.

## 2.3 | Genome assembly and quality control

To obtain high-quality genome, WGS data were used for genome survey analysis to estimate essential genome information, including genome size, level of heterozygosity, and repeat content. We used GCE pipeline (Liu et al., 2013) to estimate the genome size of *B. schroederi* before genome assembly. We generated a total of 29.32 Gb (about 110×) raw PacBio long reads. First, we used Blastall (v2.2.26) (Camacho et al., 2009) to compare the raw data with the NCBI database to confirm that the DNA of the sequenced samples was not contaminated by other species. Then we counted the length frequency distribution of TGS reads to evaluate the sequencing quality and provide a reference for the setting of subsequent assembly parameters.

The *B. schroederi* genome was assembled using a “correct-then-assemble” strategy. First, NextDenovo (v2.0-beta.1; <https://github.com/Nextomics/NextDenovo>) was used to correct and assemble a draft genome. Arrow algorithm was then used to carry out a second round of correction for this assembly. NextPolish (v1.0.5) (J. Hu, Fan, Sun, & Liu, 2020) was further used for genome polishing by using the WGS data. We then obtained a primary genome assembly with 639 contigs with N50 of 1.27 Mb. To finally ligate the scaffolds to chromosomes, Hi-C technology (Lieberman-Aiden et al., 2009) was used to capture the chromosome conformations. 105Gb (~400 X) Hi-C sequencing data were generated from a single Hi-C library which was constructed as previously described. All Hi-C reads were first mapped against our assembled genome using BWA (v0.7.13-r1126) (H. Li & Durbin, 2010) with parameters “bwa mem -t 16 -k 19 -a -V”. HiC-Pro pipeline (Servant et al., 2015) was then used for filtering the mapping result, leaving 645 million valid read pairs. Next, Juicer v1.5.7 software was used for auxiliary assembly: 1) both duplicates and near-duplicates are removed; 2) read pairs that aligned to three or more locations are set aside (Durand, Shamim, et al., 2016). Then, 3D-DNA, a custom computational pipeline, was applied for correcting misassemblies, anchor, order and orient fragments of DNA (Dudchenko et al., 2017). Finally, files generated from the 3D-DNA were loaded into the visual software Juicebox Assembly Tools module v1.11.08 for correction and review (Dudchenko et al., 2018; Durand, Robinson, et al., 2016). Contigs from the *B. schroederi* were successfully clustered into 21 groups, which were further ordered and oriented into Pseudochromosomes (Figure S2). The completeness of the genome were evaluated using sets of BUSCO with genome mode and lineage data from nematode and eukaryote, respectively (Simão, Waterhouse, Panagiotis, Kriventseva, & Zdobnov, 2015).



## 2.4 | Detection and classification of repetitive elements

For *de novo* identification of repeat elements, we constructed a transposable element (TE) library of the *B. schroederi* genome to identified types of repeat elements by using Tandem Repeat Finder (TRF)(Benson & G.), LTR.FINDER(X. Zhao & Hao, 2007) and RepeatModeler (v1.0.8)(Smit, Hubley, & Green, 2015). RepeatMasker(Tarailo-Graovac & Chen, 2009) and RepeatProteinMask(Tempel, 2012) were used to search the genome sequences for known repeat elements, with the genome sequences used as queries against the Repbase database(Jurka et al.).

## 2.5 | Protein-coding genes prediction and function annotation

A combined strategy of *de novo* gene prediction, homology-based search and RNA sequencing-aided annotation were used to perform gene prediction. For homology-based annotation, we selected the protein-coding sequences of five homologous species (*Brugia malayi*, *C. elegans*, *Pristionchus pacificus*, *Steinernema carpocapsae* and *T. canis*) from NCBI (<https://www.ncbi.nlm.nih.gov/>). For RNA-based prediction, a male and a female transcriptome sequence was aligned to the genome for assembly using TopHat (v2.1.0)(Trapnell, Pachter, & Salzberg) plus Trinity (v2.0.6)(Haas et al.) strategy. PASA pipeline (v.2.1.0) was applied to predict gene structure after which the inferred gene structures were used in AUGUSTUS (v.3.2.3)(Mario et al., 2006) to train gene models based on transcript evidence. In addition, genome sequence was analyzed by the program GeneMark (v1.0)(John & Mark, 2005) utilizing unsupervised training to build a hidden Markov model. The consistent gene sets were generated by combining all above evidence using MAKER (v.2.31.8)(Campbell, Law, Holt, Stein, & Yandell, 2013). All gene evidence was merged to form a comprehensive and non-redundant gene set using EvidenceModeler (v1.1.1, EVM)(Haas et al., 2008).

In order to perform gene functional annotation, we aligned above gene sets against several known databases, including SwissProt, TrEMBL, KEGG, COG and NR. GO information was obtained through Blast2go (v.2.5.0)(Conesa et al., 2005). Furthermore, the mitochondrial genome was assembled by blasting with *B. schroederi*'s mtDNA sequence from NCBI database(NC\_015927.1)(Xie et al., 2011). The mitochondrial genome was annotated on GeSeq online (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) using homologous gene alignment(Michael et al., 2017). Four types of Non-coding RNA (ncRNA; including tRNA, snRNA, miRNA, and rRNA) were predicted. tRNAscan-SE (v1.3.1)(Lowe & Eddy, 1997) were used to predict tRNAs. We aligned *B. schroederi* genome against Rfam (v12.0)(Kalvari et al., 2018) database and invertebrate rRNA database to predict snRNA, miRNA and rRNA, respectively.

## 2.6 | Proteases, protease inhibitors (PI) and Excretory/secretory proteins (EPs)

Proteases and PIs were annotated using MEROPS batch-BLAST(Rawlings & Morton, 2008) against MEROPS ( $E < 1e^{-10}$ ). The conventional secretion proteins and non-conventional secretion proteins were identified. SignalP-5.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) and TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) were used to predict the signal peptides and transmembrane regions of proteins, respectively. The proteins with NNScore  $> 0.9$  were selected as non-conventional protein secretion in the results of SecretomeP 2.0 Server (<http://www.cbs.dtu.dk/services/SecretomeP/>) identification.

## 2.7 | Reconstruction of phylogeny and evolutionary analysis among genomes from six nematodes

Genomes and annotation files of three roundworms parasitic on mammals (*A. suum*, *T. canis* and *P. univalens*), one free-living nematode (*C. elegans*) and one nematode parasitic on the root of a plant (*M. hapla*) were download from NCBI or WormBase database (Table S10)(Michelle, Dubaj, Price, Daryl, & Hurd, 2019). The syntenic analysis both at whole-genome nucleotide-level and protein-level were performed by aligning 5 nematodes genomes to our assembled *B. schroederi* genome by using MCScanX(Y. Wang et al., 2012) software. Orthogroups among six nematodes are define using TreeFam (v4.0)(H. Li et al., 2006). Next, the protein sequences from each family were aligned using MUSCLE (v3.8.31)(Edgar, 2004) with default parameters. The conserved CDS alignments were extracted by Gblocks(Gerard & Jose, 2007). We select single-copy gene families to construct phylogenetic trees based on maximum likelihood using

RAxML (v8.2.4)(Alexandros, 2014) with PROTCATGTR nucleotide substitution model with 500 bootstrap replicates.

Using the divergent time between *C. elegans* and *A. suum*, which was calculated based on fossil evidence, as the reference time points (Mcgill, Fitzpatrick, Pisani, & Burnell, 2017), we estimated the divergent time between each species by MCMCtree from the PAML (v4.8)(Bo & Yang, 2013) package with default parameters.

## 2.8 | Expanded and contracted gene families

Based on the phylogenetic tree we constructed using the 2451 single-copy genes, we explored significant expanded and contracted gene families in six nematodes (*A. suum*, *B. schroederi*, *T. canis*, *P. univalens*, *C. elegans* and *M. hapla*). Gene family expansion and contraction analyses were performed using CAFÉ (v4.2)(Bie, Cristianini, Demuth, & Hahn, 2006). We used a median gene number to estimate the changes in gene family size. The overall P-value of each gene family were calculated and the details of each branch and node with the exact P-values of each significant overall P-value ([?] 0.01) gene family were also calculated. Here we performed analysis in two levels: 1) Comparing gene families of roundworms to other nematode' to determine the expanded and contracted genes in roundworms; 2) Comparing the *B. schroederi* and other three roundworms (*A. suum*, *P. univalens* and *T. canis*) to identify gain and loss of genes in *B. schroederi*, and then further explain their differentiation within the roundworm branch.

## 2.9 | Identification of positively selected genes (PSGs)

Based on gene families retrieved from the TreeFam, we identify PSGs between *B. schroederi* and other five species (*A. suum*, *T. canis*, *P. univalens*, *C. elegans* and *M. hapla*). The branch-site model of CodeML in PAML (v4.8)(Bo & Yang, 2013) was used to detect potential PSGs. Using the null hypothesis and the alternative hypothesis to estimate whether the dN/dS ( $\omega$ ) value of the foreground branch (Marked as *B. schroederi*) was larger than 1 or not. Then performing likelihood ratio test (LRT) using R "chisq.test()" function to calculate chi-squared distributions with 1 degree of freedom. The PSG was identified by meeting the requirements of a corrected P-value ( $< 0.05$ ) and contained at least one positively selected site with a posterior probability [?]1.

## 2.10 | Identification of detoxification-related gene families

Seven detoxification-related gene families of four roundworms were identified by the HMMER3 (<http://hmmer.janelia.org/>) software, including ABC domain (PF00005), Cytochrome P450 (*CYP*) domain (PF00067), GST domain (PF00045), CHIA domain (PF00704), PTCHD domain (PF02460) and PTP domain (PF00782). The domain file is used as first template for scanning the gene family, the output genes were filtered out with E-value lower than  $1e^{-10}$ , then the filtered genes were used as second template for scanning the gene family. Again, the second output genes were filtered out with E-value lower than  $1e^{-10}$ .

## 2.11 | The change of effective population size on the function of time

We inferred the demographic history of the *B. schroederi* by use of WGS data generated by DNBSEQ-T1 from one individual. We simultaneously perform the same analysis on a giant panda by using resequencing short reads of an individual download from SRA database (accession SRA053353). For this analysis, we used BWA (v0.7.13-r1126)(H. Li & Durbin, 2009) to map the clean reads to each genome with the default parameters. Next, the PSMC method(H. Li & Durbin, 2011) was used to evaluate dynamic change of effective population size ( $N_e$ ) of *B. schroederi* and giant panda. Following Li's procedure(H. Li & Durbin, 2011), we applied a bootstrapping approach, repeat sampling 100 times to estimate the variance of simulated results for both *B. schroederi* and giant panda. We used 0.17 and 12 years per generation (g) and mutation rate ( $\mu$ ) of  $9 \times 10^{-9}$  and  $1.29 \times 10^{-8}$  for *c* and giant panda, respectively(Cutter, 2008). Since fluctuations in the effective population size of giant pandas have been reported to closely reflect changes in climate and atmospheric dust(S. Zhao et al., 2013), we added the mass accumulation rate (MAR) of Chinese loess over the past 250,000 years for comparison. In addition, we implemented the MSMC2(Schiffels & Durbin, 2014) which can infer the recent effective population size history. We phased all SNPs of each individual by using

beagle (v5.0)(Browning & Browning, 2007), then it calculated used the following parameters: -i 20 -t 6 -p '10\*1+15\*2'. The mutation rate ( $\mu$ ) of *B. schroederi* for MSMC2 were used the same values as for PSMC.

## 2.12 | Population structure of *B. schroederi* in Qinling and Sichuan

A total of 240 samples collected from individuals in captivity and 26 samples from individuals in wild were re-sequenced using the DNBSEQ-T1&T5 platform. High-quality reads were aligned to the reference genome using BWA-MEM (0.7.13-r1126)(H. Li & Durbin, 2010) with default parameters. SAMtools (v0.1.19)(H. Li et al., 2009) and Genome Analysis Toolkit (GATK v 4.0.3.0)(Depristo, Banks, Poplin, Garimella, & Daly, 2011) were used to obtain the SNP set within the population. Hard filtering was applied to the raw variant set using "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" -filter-name "snp\_filter". SNPs with >0.5% missing data or <0.01 minor allele frequency (MAF) were filtered out using vcftools (v0.1.12a)(Danecek et al., 2011). PCA analysis of SNPs was carried out using EIGENSOFT(Nick et al., 2006) software, and the population clustering analysis was conducted in PLINK(Purcell et al., 2007). We used the whole-genome SNPs to construct the ML phylogenetic tree with 1000 bootstrap using iqtree (v1.6.12)(Lam-Tung, Schmidt, Arndt, & Quang, 2015), and using an genome sequence information of *P. univalens* as an outgroup. Population structure of all 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 K from 2 to 5 and ran the cross-validation error (CV) procedure.

## 2.13 | Recent nature selection analysis

Extended Haplotype Homozygosity (EHH) and iHS methods was used for detecting SNPs under strong positive selection of Captive and wild population, respectively(Mathieu & Renaud, 2012). We searched for genes in the 5-kb flanking region from both sides of candidate SNPs, and calculate the accumulated iHS scores by adding each iHS score of the top 0.5% SNPs. Next, to uncover genetic variants involved in adaptation of different environments, we performed comparisons between Captive and wild population. The XP-EHH method was used to detect selective sweeps using the R package rehh (v3.1.2;<https://cran.r-project.org/web/packages/rehh/vignettes/rehh.html>). We then split the genome into non-overlapping segments of 50 kb to use the maximum (positive) XP-EHH score of all SNPs. The regions with P values less than 0.01 were considered significant signals in the population of interest.

## 2.14 | Known and potential drug targets

All compound-related proteins were searched against target proteins from ChEMBL v26(Anna et al.) using BLASTP (E[?] $1 \times 10^{-10}$ ). We screened out all types of single proteins in the ChEMBL database for blasting. Known drug target were identified from available publications and by searching for 'anthelmintics' in the DrugBank(S et al., 2017) database. For potential drug targets, we screen out all the single protein targets as described(Coghlan, Mutowo, O'Boyle, Lomax, & Berriman, 2018) and made adjustments to evaluate each gene. We set a score of '0/1' considering six main factors: 1. Similarity with ChEMBL drug targets and a highly conserved alignment (>80%); 2. Lack of human homologues; 3. Related to lethal, L3 arrest, flaccid, molt defect and sterile phenotypes. Lethal phenotypes were identified in WormBase WS240; 4. Whether the protein was a predicted chokepoint enzyme(Tyagi, Seshadri, Parkinson, & Mitreva, 2018); 5. The protein was predicted as an excretory/secretory protein (EP); (6) The protein had a structure in the PDBe(Sameer et al., 2016). In order to make the screening most efficient, we searched for commercially available compounds against the target protein in ZINC15(Sterling & Irwin, 2015). Finally, we selected compounds approved in phase III or above as suggested chemical compounds.

# 3 | RESULTS

## 3.1 | Genome assembly, annotation and evaluation

A total of 29 Gb (110 X) PacBio long reads were generated (Table S1). The genome size was estimated to be 266.8 Mb based on K-mer depth distribution analysis (Tables S2 and Fig. S1), and the size of the assembled *B. schroederi* genome reached 262 Mb, accounting for 98.48% of the estimated genome. This genome contained 106 scaffolds, 21 of which were superscaffolds ligated using 105 Gb (~386 X) Hi-C sequencing data (Fig. 1a;

Fig. S2). The total length of these 21 superscaffolds reached ~263 Mb, accounting for 98.67% of the whole genome (Table S3). The final scaffold N50 was 12.69 Mb (Table 1), which is significantly better than the published genome (Y. Hu et al., 2020). The GC-depth distribution (Fig. S3a and S3b) further showed that most genomic regions have a GC content narrowly centered around 37%, which is similar to that of other three roundworms ("Ascaris suum draft genome," 2011; Zhu et al., 2015) (*Ascaris suum*, *Parascaris univalens* and *Toxocara canis*; Table 1). Benchmarking Universal Single-Copy Orthologs (BUSCO) scores against nematode and eukaryote databases were 91.9% and 93.4%, respectively (Fig. S4), reflecting the highest genome completeness among published roundworm genomes (Table 1). 19,291 protein-coding genes were predicted via *ab initio*, homology-based and RNA sequencing-aided methods (see Methods) ("serine-threonine kinase KIN-29 modulates TGFbeta signaling and regulates body size formation,"). Kyoto encyclopedia of genes and genomes (KEGG), clusters of orthologous groups of proteins (COG), TrEMBL, gene ontology (GO), Swissprot and InterPro (Fig. S5 and Fig. S6). The average length of coding sequences (CDS) was 1,052 bp with an average of 6.87 exons per gene, which is similar to that of other related roundworms (Table S4). Furthermore, a 14,767 bp mitochondrial genome was identified, containing 12 genes encoding proteins, 2 genes encoding rRNAs, and 22 genes encoding tRNAs (see methods; Table S5 and Fig. S7). To evaluate the completeness of the predicted protein-coding genes, we compared the length distributions of mRNA, CDS, exons and introns in *B. schroederi* with those in other five nematodes (Fig. S8 and Fig. S9).

Total repeats (DNA transposons and RNA transposons) accounted for 11.76% of the genome (Table S6-S8 and Fig. S10). Huge variation in the proportion of repeat content is found among published nematode genomes (from 1% to 48%) (Berriman et al., 2009; International, Helminth, Genomes, & Consortium, 2018; Schiffer, Kroiher, Kraus, Koutsovoulos, & Schierenberg, 2013). Transposable elements (TEs) account for 9.51% of the *B. schroederi* genome (Table S8), while TEs constitute 4.4% and 13.5% in the genomes of *A. suum* ("Ascaris suum draft genome," 2011) and *T. canis* (Zhu et al., 2015), respectively. We identified a significant expansion of DNA transposons in *B. schroederi* compared to *T. canis* (Zhu et al., 2015) and *A. suum* ("Ascaris suum draft genome," 2011) (Supplementary Data 1). There are at least 64 DNA transposon families of which CMC-EnSpm, DNA and MULE-MuDR dominated the genome. We identified 17 long terminal repeats (LTRs) retrotransposon and 41 non-LTRs retrotransposon families (25 LINE and 16 SINE). Pao and Gypsy are the predominant LTRs, and CR1, RTE-RTE and L2 are the predominant non-LTRs. The number and size of the retrotransposon families were similar to those of other related roundworms ("Ascaris suum draft genome," 2011; Ghedin et al., 2007; Zhu et al., 2015).

### 3.2 | Development-related genes for key enzymes, ion channels, receptors and secretome

To understand the genetic basis for the adaptation of *B. schroederi* to the parasitic life, we assessed the abundance of several major protein classes in *B. schroederi*, *A. suum*, *P. univalens*, and *T. canis* (Table S10 and Fig. S11b). A genome-wide search enabled us to identify 62 G-protein coupled receptors (GPCRs), 437 proteases and protease inhibitors (Supplementary Data 2). Some chemoreceptor families, especially those differentially expressed during the life cycle, were almost completely conserved among nematodes. For example, the homologues of *Caenorhabditis elegans* *daf-37* (GeneID: Baysch11898) and *daf-38* (GeneID: Baysch06944), which are known to mediate ascarioside signaling, are expressed during the transition from dauer larvae to infective larvae (Park et al., 2012). By comparing with the ligand-gated ion channel gene set collected from wormBase, 65 genes were identified, including members of the previously described nematode acetylcholine receptor classes (*deg-3*, *acr-16*, *unc-29* and *unc-38*) (International et al., 2018). We predicted excretory/secretory proteins (EPs) of *B. schroederi*, and at least 1,395 EPs (5.26% of total protein) with diverse functions were identified, including 1,046 conventionally secreted proteins and 349 nonconventionally secreted proteins (Supplementary Data 2).

### 3.3 | Evolutionary and phylogenetic relationships among six nematodes

We clustered the *B. schroederi* gene models with the genes from five other nematode genomes (*A. suum*, *P. univalens*, *T. canis*, *C. elegans*, and *Meloidogyne hapla*; Table S11). We found that the six nematodes share 3,906 homologous gene families (Fig. 1c). In addition, four roundworms show high consistency in the number of single-copy and multi-copy genes (Fig. S11c). Collinearity results showed that although several

roundworms are closely related, collinearities among the genomes are low (Fig. 1b). The proportions of collinearity between *B. schroederi* and the other three roundworms in the genome species are 35.37% (vs *T. canis*), 43.86% (vs *A. suum*), and 55.12% (vs *P. univalens*), respectively, which indicate that genetic differentiation among roundworms is considerable.

We used 2,451 single-copy genes shared within the six nematode genomes to reconstruct a phylogenetic tree (Fig. 1d). The relationships among the six nematodes in the phylogenetic tree are consistent with a previous study (International et al., 2018). *B. schroederi* is closely related to *A. suum* and *P. univalens*. According to the TimeTree (Hedges, 2011) database and fossil evidence from *A. suum* and *C. elegans* (McGill et al., 2017), we estimated the divergence time to approximately 400-269 million years ago (Mya), and the divergence time between the four roundworms is approximately 160-26 Mya (Fig. 1d). Among the four roundworms, *T. canis* was identified as the earliest branch to the other three roundworms (approximately 134 Mya).

### 3.4 | Expanded and contracted genes in Ascariasis

Compared with *C. elegans* and *M. hapla*, a large number of genes have been lost or contracted in the branch of roundworms (Fig. 2a). Specifically, we found that 563 gene families are significantly contracted ( $P < 0.05$ ), with an average loss of 1.61 genes in each family (Table S12). However, roundworms also show significant expansion in 29 gene families ( $P < 0.05$ ). We focused on the changes in gene number related to free life and those related to parasitic life in nematodes. For both the expanded and contracted gene families, KEGG enrichment analysis showed that pathways related to tissue development, metabolism and environmental information processing had undergone significant changes (Fig. S12a and S12b). Interestingly, a similar expansion also appeared in the roundworm branch compared to *M. hapla* (Fig. S12c-d). For tissue development, we found that the *CPG-2* gene family is significantly expanded in the roundworm branch ( $P < 0.01$ ). In addition, a significant expansion of the tight junctions (ko04530), phagosome pathway (ko04145) and Rap1 signaling pathway (ko04015) was observed ( $P < 0.01$ ; Fig. 2a). In relation to self-defense, consistent with a previous study, we observed an expansion of the chymotrypsin/elastase inhibitor gene family, which may be related to a protection of roundworms from host proteases (International et al., 2018). In addition, according to the copy number statistics of expanded gene families, the actin family was significantly expanded among all four roundworms (Fig. 2b). GO enrichment analysis showed that the genes related to nematode behavior and biological adhesion accounted for the most significant difference among all gene families exhibiting expansions in roundworms (Fig. 2c).

### 3.5 | Expanded or contracted gene families related to the adaptation of *B. schroederi*

To better understand the adaptation to the specific habitat and intestinal environment of the giant panda, we analyzed the expanded and contracted genes in *B. schroederi* compared with three roundworms (*A. suum*, *P. univalens* and *T. canis*).

We identified expanded gene families with functions involved in striated muscle contraction (GO:0006941), nematode larval development (GO:0002119), larval feeding behavior (GO:0030536), chitin metabolic process (GO:0006030) and actin cytoskeleton organization (GO:0030036). The most significantly enriched GO term was the straight muscle contraction, which was largely due to the highly significant expansion of the actin family (Fig. 3a, 3d). KEGG enrichment showed that the number of genes involved in metabolic pathways, including drug metabolism (metabolism of xenobiotics by cytochrome P450, ko00980;  $P < 0.01$ ) and autoimmunity (ko05130, ko05100;  $P < 0.01$ ), exhibited significant changes (Fig. 3a). We observed an expansion of the gene family involved in positive regulation of eating behavior (Fig. 3b). Finally, we found that *B. schroederi* has 654 unique annotated proteins, which were mainly enriched in the synthesis and recycling pathways of essential amino acids especially the degradation of valine, leucine and isoleucine, (ko00280,  $P < 0.01$ ; Fig. 3c).

### 3.6 | The significant expansion of actin family in Ascariasis and positive selection genes (PSGs) in *B. schroederi*

The migration of roundworms is the main cause of VLM (Fig. 4a). Actin polymerization is controlled by

intracellular signals that are mediated by small GTPases of the Rho family (*RhoA*, *Rac1*, and *Cdc42*; Fig. 4c). We observed a very significant expansion of the actin family in the *B. schroederi* genome (Fig. 3d). Surprisingly, the three upstream regulators of actin (*Rac1*, *ROCK* and *MLCK*), were under strong positive selection ( $P < 0.01$ ). *ROCK* phosphorylates the LIM-kinase which then phosphorylates cofilin to promote rho-induced actin cytoskeleton reorganization (Maekawa & M., 1999). These findings suggest a possible effect on myosin-actin interaction. Using the branch-site model implemented in PAML, 475 genes in the *B. schroederi* genome were found to be under strong positive selection compared with other three roundworms (Supplementary Data 3a). Compared with three roundworms, the acetylcholine receptor subunit alpha-type deg-3 ( $P = 6.5 \times 10^{-3}$ ), which is an important drug target (Jones, Davis, Hodgkin, & Sattelle, 2007), was shown to be under strong positive selection in the *B. schroederi* genome (Supplementary Data 3a).

### 3.7 | Demographic history of *B. schroederi*

To reconstruct the demographic history of *B. schroederi*, and explore the relationship with the giant panda, we used the pairwise sequentially Markovian coalescent (PSMC) model to estimate the changes of the effective population size of both *B. schroederi* and the giant panda within the last one million years (H. Li & Durbin, 2011). PSMC analysis showed that the effective population sizes ( $N_e$ ) of *B. schroederi* and the host giant panda have almost exactly the same trend from 300 Kya (thousand years ago) to 10 Kya (Fig. 5a), but the change of effective population size of the roundworms showed a slight lag. The effective population size of the giant panda declined significantly during the last two Pleistocene glacial periods (300-130 Kya and 50-10 Kya), and the effective population size of the roundworm also reached a historical low level. The most obvious change occurred in the Greatest Lake Period (50-30 Kya), where the effective population size of the two species reached their pinnacle (Jinchu & Wei, 2004). In addition, we applied Multiple Sequentially Markovian Coalescent (MSMC) method (Schiffels & Durbin, 2014) to infer the recent demographic events of *B. schroederi* and performed 5 repetitions (each repetition uses 4 individuals per population). The effective population size of roundworms showed a sharp decline in the last 10,000 years, which also may be related to the host population dynamics. Studies have indicated that human activities may have caused the decline of the giant panda population in the past few thousand years, and the roundworm population may be affected by this.

### 3.8 | Population structure of *B. schroederi* populations

We carried out whole-genome resequencing of 266 samples, including 240 from captive pandas of the Sichuan subspecies and 26 samples from individuals of the wild Qinling panda subspecies (Fig. 6a). The average sequencing coverage and sequencing depth reached 97.91% and 41-fold, respectively (Table S1 and S13). A total of 6.32 million SNPs were obtained after filtering (see methods). Principal components analysis (PCA) supported the clear separation between *B. schroederi* from the captive Sichuan and the wild Qinling panda subspecies (Fig. 6b), with PC1 separating the Qinling and Sichuan populations and PC2 separating the Sichuan population into two clusters ( $P < 0.05$ ). We constructed a phylogenetic tree using the maximum likelihood (ML) method, which showed two distinct clusters in the whole population, with all Qinling individuals forming a single cluster and all Sichuan individuals forming another single cluster (Fig. 6c). In addition, the results of structure analyses also indicated that there were almost no shared ancestral components between the Qinling and Sichuan populations, supporting the results from the phylogenetic tree and the PCA (Fig. 6d). Although the two populations showed extremely similar genetic diversity (Fig. S13 and Fig. S14), our results consistently supported two distinct groups corresponding to the Sichuan and Qinling populations.

### 3.9 | Recent positive selections in the *B. schroederi* populations

We used integrated haplotype score (iHS) to detect genes under recent natural selection in the captive and wild populations. A total of 29,553 SNPs in captive and 18,953 SNPs in wild were identified within the top 1% iHS scores. By extending the 25 kb distance around the top 1% SNPs, filtering out SNPs in the non-gene regions, a total of 518 and 370 genes were located in the positively selected regions in captive and wild populations, respectively (Supplementary Data 3b and 3c). We further calculated the distribution

of nucleotide diversity on the 21 superscaffolds (Fig S13), and found that the genetic diversity in regions around the glutamate-gated chloride channel alpha (*glc-1* , a receptor for anthelmintic ivermectin(Cook et al., 2006)), nose resistant to fluoxetine protein 6 (*nrf-6* , a fluoxetine (Prozac) resistance gene(Choy, Kemner, & Thomas, 2006; Fares & Grant, 2002)), ABC transporter ced-7 (*ced-7* , phagocyte corpse)(Wu & Horvitz, 1998) and  $\beta$ -1,4-N-acetylgalactosaminyltransferase bre-4 (*bre-4* , resistance to pore-forming toxin(Griffitts et al., 2003)) genes were significantly lower than in the flanking genome regions (Fig. 7b). Interestingly, the four genes were only found under positive selection in the captive population, but not in the wild population. In addition, we also used the cross-population extended haplotype homozygosity (XP-EHH) method(Pardis C Sabeti et al., 2007) to screen for genes that may have been positively selected by different deworming selection pressures by comparing the captive and wild populations (Supplementary Data 3d). Similarly, the genes encoding the multidrug resistance protein *pgp-3* (Xu et al., 1998) which is related to ivermectin resistance, *glc-1* , *nrf-6* , cytochrome p450 (*CYP* ) family members and other drug resistance related genes, were also observed to be under strong positive selection in the captive population (Fig. 7c).

### 3.10 | Identifying anthelmintic resistance-related gene family and drug targets

We use the HMMER3 software to scan several detoxification-related gene families at the whole genome scale, including ATP dependent (*ABC* ), cytochrome P450 (*CYP*), glutathione s-transferase (*GST*), glycoside hydrolase family 18 (*CHIA*), patched family (*PTCHD*) and protein tyrosine phosphatase family (*PTP*) (Fig S11d). A total of 97 *ABC* transporters, multipass membrane proteins, were identified in *B. schroederi* , and the average number of *ABC* transporter genes in roundworms was greater than that in *C. elegans* (60)(Schumacher & Benndorf, 2017). The high-quality genome data of *B. schroederi* provides an opportunity to identify biologically active anthelmintic compounds. On the one hand, it enables identification of the targets of existing anthelmintics, on the other hand, it also enables identification of new potential targets for compounds from other areas of drug discovery. All compound related proteins were searched against target proteins from ChEMBL v26(Anna et al.) using BLASTP ( $E[?]1 \times 10^{-10}$ ), and a total of 4,554 small molecules with recorded biological activities were identified. By blasting against the ChEMBL(Anna et al.) databases, a total of 90 known genes, which encode specific drug targets were identified. The corresponding drugs (13 drugs used to treat humans with World Health Organization (WHO) ATC code P02 ‘WHO anthelmintics’ and 10 drugs from DrugBank(S et al., 2017)) were further collated by searching DrugBank databases and the literature (Supplementary Data 4a). Some of these drugs have been proven to be effective against *B. schroederi* , such as albendazole(Fu et al., 2011), mebendazole(Bourne, Cracknell, & Bacon, 2010), pyrantel(Xie et al.) and ivermectin(C. Wang et al., 2015). Many existing anthelmintics are compromised by the increase of resistance in roundworm populations(D. Li et al., 2015). In addition to known drugs, we were committed to identifying new potential drug targets. We focused on single protein ChEMBL targets that may be easier to develop drugs against than protein complexes(International et al., 2018). By blasting against target proteins (similarities > 80%) in the single protein database ChEMBL, we identified 95 genes encoding single proteins. Then we set a score of ‘0/1’ considering six main factors to evaluate the potential of the protein as a drug target (see methods; Fig. 8). Finally, we located the position of all the drug target encoding genes on superscaffolds (Fig. 8). Since the existing Phase III and above drugs have greater potential for development into new anthelmintics, we searched for commercially available compounds against each target protein although these compounds were not originally designed as anthelmintics. Among all the proteins, we found that three target genes (*cmd-1* , *Ap2s1* , *HRAS* ) have available compounds with Phase III/IV approvals (Supplementary Data 4b). These potential drug targets and compounds will provide references for the development of new anthelmintics.

## 4 | DISCUSSION

*B. schroederi* exhibits strong environmental adaptability and wide distribution, and is a threat to the health of giant pandas (Zou et al., 1998). The in-depth studies of *B. schroederi* have been hampered by the lack of a high-quality genome. The scaffold N50s of published *A. suum* , *P. univalens* and *T. canis* genomes are 290 kb, 1,825 kb and 375 kb, respectively (Table 1). In this study, we present the first chromosome-scale genome assembly of the *B. schroederi* with the scaffold N50 of 12.69 Mb, representing a genome assembly

with the best contiguity in Ascarididae. We envisage that this genome will provide a valuable and useful genetic resource for future research on roundworms, as well as drug development for expulsion.

Roundworms have special characteristics that are different from free-living nematodes reflecting the adaptation to the parasitic life. Eggs of roundworms have a tough and elastic polysaccharide chitin shell, which enables eggs to persist in the soil for up to ten years (FAIRBAIRN, 1970). We have observed a significant expansion of the chitin-binding protein *CPG-2* family in roundworm branches, which may be related to the formation of the roundworm egg shell, thereby prolonging the survival of roundworms even in a harsh environment. In addition, in the parasitic stage, larvae enter the intestine, penetrate the intestinal wall, migrate among tissues and organs (K. Kazacos & W. M. Boyce, 1989), molt and develop, finally return to the small intestine to develop into adults and lay eggs again. Some genes potentially involved in tissue invasion and immune evasion have been significantly expanded in roundworms, including genes homologous to metalloproteinase and serine/threonine-protein kinase, respectively. Previous studies have shown that metalloproteinase(s) in the secretory products of astacins in the nematode epidermis can digest collagen in host tissues, and thus be involved in the migration of larvae in viscera (Hanns et al., 2011; Williamson et al.).

Although the morphological characteristics of Ascarididae roundworms are similar, the *B. schroederi* still shows unique molecular evolutionary traits. The giant panda has gradually evolved in response to the bamboo diet during millions of years of evolution (Zhou, Hu, Yuan, & Wei, 1997). However, the giant panda has maintained the intestinal structure of carnivores, with short and thick small intestines (Guibo Yang, 1995). Due to the low digestibility of bamboo, giant pandas have to ingest large amounts of bamboo to meet their nutrient needs (Sims et al., 2007), and a large amount of feces is produced and discharged every day. Accordingly, *B. schroederi* needs to absorb nutrients as much as possible. In the *B. schroederi* genome, several unique gene families of *B. schroederi* were found to be involved in the metabolism of essential amino acids, especially the degradation of valine, leucine and isoleucine (KO00280;  $P < 0.01$ ), which is likely to enhance the ability of *B. schroederi* to absorb nutrients. In addition, *B. schroederi* needs stronger motor ability than *A. suum*, *P. univalens* and *T. canis* to survive in the small intestine because of its much smaller body size, and at the same time ensure that they can avoid expulsion. The muscle tissue of the roundworm plays a key role in motility and the extreme expansion and positive selection of the actin family may have provided the driving force for muscle contraction and cell movement (Hall, 1998). Actin promotes muscle contraction and plays a very important role in the movement and migration of *B. schroederi* in the host. Studies have shown that actin is involved in the repair of nematode epidermis damage (Suhong & Chisholm, 2012), which is of great significance to the migration of *B. schroederi* in the giant panda. The expansion of the actin gene family may, at least to some extent, explain the genetic basis of stronger locomotion ability of *B. schroederi* than other roundworms.

According to a previous investigation, the cause of death of giant pandas in recent decades has shifted from starvation and poaching to VLM-related deaths (J.-S. Zhang et al., 2008). Frequent use of drugs may drive the increasing frequency of genes related to drug resistance in the population, leading to widespread drug resistance in the *B. schroederi* population. Furthermore, there have been reports of side effects in giant pandas after the administration of existing anthelmintic drugs (C. Wang et al., 2015). We identified the *ABC*, *CYP*, *GST*, *CHIA*, *PTCHD* and *PTP* gene families in the *B. schroederi* genome. These genes may be involved in the metabolism of drugs and other xenobiotics and/or biosynthesis and metabolism of endogenous compounds. We observed a recent significant positive selection of *ABC* and *CYP* family members and other resistance-related genes (*glc-1*, *nrf-6*, *pqp-3* and *bre-4*) in captive (SC) populations. Although wild and captive populations were obtained from two different regions (Qinling and Sichuan), natural selection analysis mainly considers recent changes in gene frequency. The two populations are facing completely different selection pressures for deworming, and thus, offer an option for evaluating natural selection trends of a few resistance-related genes. The results indicated an increased frequency of drug resistance-related genes in captive populations. This may be related to the frequent use of drugs in recent decades. Although the degree of natural selection in the current resistance areas cannot be quantified, it is possible that the gene frequency of these genes is still increasing, and it may cause the emergence and increase of resistant individuals. Studies have shown that some new sources of infection may even evolve into



potential antibiotic-resistant pathogens(Zumla & Hui, 2019). Therefore, the identification of drug-resistance genes and the detection of drug-resistant individuals are still essential in future works.

There is an urgent need for new anthelmintic drugs for intestinal expulsion of roundworms(James, Hudson, & Davey, 2009; Jia, Melville, Utzinger, King, & Zhou, 2012). Specifically, there is a pressing need for new anthelmintic drugs to protect the giant panda, since existing drugs suffer from low efficacy, serious side effects or rising drug resistance in parasite populations due to increased frequency of use(C. Wang et al., 2015) . The chromosome-scale genome of *B. schroederi* provides a reference for the development of species-specific drugs, and drug targets can be screened from the whole genome level. We identified a total of 90 known drug targets and 95 potential drug targets, providing a basis for the development of follow-up drugs and vaccines. We searched four compounds (lonafarnib, haloperidol, trifluoperazine and chlorpromazine; Supplementary Data 4b) that have a phase 3/4 approval. These compounds could be considered for repurposing as novel anthelmintics, which would save considerable effort and expense. Nevertheless, the anthelmintic activity of these compounds and other potential target compounds needs further testing. We envision that such works will provide new modalities for the prevention and treatment of baylisascariasis and other parasitic diseases.

## 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., and H.L. designed and initiated the project. Z.W.P., D.S.L., S.W.H., Y.L.W., H.T.S., H.L.C., and L.H.D. collected the samples. X.F.C., L.H., L.L.L., Y.X.L., Y.Q.Z., J.Y.Y. and H.R.L. performed the DNA extraction, library construction and sequencing. L.H., T.M.L., H.M.L., R.B.H., Q.W., Y.X.Z., and S.C.Y. performed the data analysis. L.H., T.M.L. and wrote the manuscript. Z.J.H., K.K., M.L., D.S.L., H.L., Q.L., H.M.Y. and X.X. 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., K.K., M.L., D.S.L., H.L., Q.L., H.M.Y. and X.X. Z.J.H., H.L., K.K. and Q.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)113 of China National GeneBank DataBase (CNGBdb)114 with CNSA project ID CNP0001242.

## REFERENCES

- Alexandros, S. (2014). RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (9), 9.
- Anna, G., Bellis, L. J., Patricia, B. A., Jon, C., Mark, D., Anne, H., . . . Bissan, A. L. ChEMBL: a large-scale bioactivity database for drug discovery.
- Ascaris suum draft genome. (2011). *Nature*, 479 (7374), 529-533.
- Benson, & G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27 (2), 573-580.
- Berriman, M., Haas, B. J., Loverde, P. T., Wilson, R. A., Dillon, G. P., & et al. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature*, 460 (7253), 352-358.
- Bethony, J., Brooker, S., Albonico, M., Geiger, S. M., Loukas, A., Diemert, D., & Hotez, P. J. (2006). Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The lancet*, 367 (9521), 1521-1532.
- Bie, T. D., Cristianini, N., Demuth, J. P., & Hahn, a. M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*, 22 (10), 1269-1271.
- Bo, X., & Yang, Z. (2013). pamlX: A Graphical User Interface for PAML. *Molecular Biology & Evolution* (12), 12.
- Bourne, D., Cracknell, J., & Bacon, H. (2010). Veterinary issues related to bears (Ursidae). *International zoo yearbook*, 44 (1), 16-32.
- 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.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *Bmc Bioinformatics*, 10 (1), 421.
- Campbell, M. S., Law, M., Holt, C., Stein, J. C., & Yandell, M. (2013). MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiology*, 164 (2), 513.
- Choy, R. K. M., Kemner, J. M., & Thomas, J. H. (2006). Fluoxetine-Resistance Genes in *Caenorhabditis elegans* Function in the Intestine and May Act in Drug Transport. *Genetics*, 172 (2), 885-892.
- Coghlan, A., Mutowo, P., O'Boyle, N., Lomax, J., & Berriman, M. (2018). Creating a screening set of potential anthelmintic compounds using ChEMBL.
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21 (18), 3674-3676.
- Cook, A., Aptel, N., Portillo, V., Siney, E., Sihota, R., Holden-Dye, L., & Wolstenholme, A. (2006). *Caenorhabditis elegans* ivermectin receptors regulate locomotor behaviour and are functional orthologues of *Haemonchus contortus* receptors. *Molecular and biochemical parasitology*, 147 (1), 118-125.
- 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.
- De Silva, N. R., Brooker, S., Hotez, P. J., Montresor, A., Engels, D., & Savioli, L. (2003). Soil-transmitted helminth infections: updating the global picture. *Trends in parasitology*, 19 (12), 547-551.

- 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.
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., . . . Aiden, A. P. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, 356 (6333), 92-95.
- Dudchenko, O., Shamim, M. S., Batra, S., Durand, N. C., Musial, N. T., Mostofa, R., . . . Stamenova, E. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *Biorxiv* , 254797.
- Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, 3 (1), 99-101.
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, 3 (1), 95-98.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* (5), 5.
- FAIRBAIRN, D. (1970). Biochemical adaptation and loss of genetic capacity in helminth parasites. *Biological Reviews*, 45 (1), 29-72.
- Fares, H., & Grant, B. (2002). Deciphering Endocytosis in *Caenorhabditis elegans*. *Traffic*, 3 (1).
- Fu, Y., Nie, H.-M., Niu, L.-L., Xie, Y., Deng, J.-B., Wang, Q., . . . Wang, S.-X. (2011). Comparative efficacy of ivermectin and levamisole for reduction of migrating and encapsulated larvae of *Baylisascaris transfuga* in mice. *The Korean journal of parasitology*, 49 (2), 145.
- 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.
- Gerard, T., & Jose, C. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56 (4), 564-577.
- Ghedini, E., Wang, S., Spiro, D., Caler, E., Qi, Z., Crabtree, J., . . . Miranda-Saavedra, D. (2007). Draft genome of the filarial nematode parasite *Brugia malayi*.
- Griffitts, J. S., Huffman, D. L., Whitacre, J. L., Barrows, B. D., Marroquin, L. D., Muller, R., . . . Aroian, R. V. (2003). Resistance to a Bacterial Toxin Is Mediated by Removal of a Conserved Glycosylation Pathway Required for Toxin-Host Interactions. *Journal of Biological Chemistry*, 278 (46), 45594-45602.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., . . . Lieber, M. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8 (8), 1494-1512.
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., . . . Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome biology*, 9 (1), R7.
- Hall, A. (1998). Rho GTPases and the Actin Cytoskeleton. *Science*, 279 (5350), p.509-514.
- Hanns, Soblik, Abuelhassan, Elshazly, Younis, Makedonka, . . . Renard. (2011). Life cycle stage-resolved proteomic analysis of the excretome/secretome from *Strongyloides ratti*—identification of stage-specific proteases. *Molecular & Cellular Proteomics Mcp* .

- Hedges, S. B. (2011). TimeTree2: species divergence times on the iPhone. *Bioinformatics*, 27 (14), p.2023-2024.
- Hotez, P. P. J., Fenwick, A., Savioli, L., & Molyneux, D. H. (2009). Rescuing the bottom billion through control of neglected tropical diseases. *Lancet*, 373 (9674), 1570-1575.
- Hu, H., Zhang, X., Pei, J., Su, L., Zhang, H., Liu, Y., & Wu, X. (2018). Investigation on the Morphology and Infection Situation of Intestinal Parasites in the Wild Giant Pandas. *Journal of Economic Animal*, 22 (2), 106-111+124.
- Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics* .
- Hu, Y., Yu, L., Fan, H., Huang, G., Wu, Q., Nie, Y., . . . Wei, F. (2020). Genomic signatures of coevolution between non-model mammals and parasitic roundworms. *Molecular Biology and Evolution* .
- International, Helminth, Genomes, & Consortium. (2018). Comparative genomics of the major parasitic worms. *Nature Genetics* .
- James, C. E., Hudson, A. L., & Davey, M. W. (2009). Drug resistance mechanisms in helminths: is it survival of the fittest? *Trends in parasitology*, 25 (7), 328-335.
- Jia, T.-W., Melville, S., Utzinger, J., King, C. H., & Zhou, X.-N. (2012). Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. *PLoS neglected tropical diseases*, 6 (5).
- Jinchu, H., & Wei, F. (2004). Comparative Ecology of Giant Pandas in the Five Mountain Ranges of Their Distribution in China.
- John, B., & Mark, B. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* (suppl\_2), suppl\_2.
- Jones, A. K., Davis, P., Hodgkin, J., & Sattelle, D. B. (2007). The nicotinic acetylcholine receptor gene family of the nematode *Caenorhabditis elegans* : an update on nomenclature. *Invertebrate Neuroscience*, 7 (2), 129-131.
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic & Genome Research*, 110 (1-4), 462-467.
- Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., & Petrov, A. I. (2018). Non-coding RNA analysis using the Rfam database. *Current protocols in bioinformatics*, 62 (1), e51.
- Kazacos, K., & Boyce, W. M. (1989). Baylisascaris larva migrans. *Journal of the American Veterinary Medical Association*, 195 (7), 894-903.
- Kazacos, K. R., & Boyce, W. M. (1989). Baylisascaris larva migrans. *J Am Vet Med Assoc*, 195 (7), 894-903.
- 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, D., He, Y., & Deng, L. (2015). Deworming Experiments of Ivermectin and Pyrantel Pamoate on Baylisascaris schroedari of Captive Giant Panda. *Animal Husbandry & Veterinary Medicine*, 47 (06), 87-90 (in Chinese).
- Li, D., He, Y., Wu, H., Wang, C., Li, C., Lan, J., . . . Yang, G. (2014). Prevalence of helminths in captive giant pandas. *Journal of Economic Animal*, 18 (4), 214-220.

- Li, H., Avril, C., Jue, R., James, C. L., Jean-Karim, H., Lara, O., . . . Lars, B. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucl Acids Research*, *34* (Database issue), D572.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* .
- 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.
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Dorschner, M. O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326* (5950), 289-293.
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., . . . Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv preprint arXiv:1308.2012* .
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25* (5), 955-964.
- Maekawa, & M. (1999). Signaling from Rho to the Actin Cytoskeleton Through Protein Kinases ROCK and LIM-kinase. *Science*, *285* (5429), 895-898.
- Mario, S., Oliver, K., Irfan, G., Alec, H., Stephan, W., & Burkhard, M. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* (suppl\_2), suppl\_2.
- 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.
- Mcgill, L. M., Fitzpatrick, D. A., Pisani, D., & Burnell, A. M. (2017). Estimation of phylogenetic divergence times in Panagrolaimidae and other nematodes using relaxed molecular clocks calibrated with insect and crustacean fossils. *Nematology*, *19* (8).
- Michael, T., Pascal, L., Tommaso, P., S., U.-J. E., Axel, F., Ralph, B., & Stephan, G. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* (W1), W1.
- Michelle, Dubaj, Price, Daryl, & Hurd. (2019). WormBase: A Model Organism Database. *Medical Reference Services Quarterly* .
- Murray, W. J. (2002). Human infections caused by the raccoon roundworm, Baylisascaris procyonis. *Clinical Microbiology Newsletter*, *24* (1), 1-7.
- Nick, Patterson, Alkes, L., Price, David, & Reich. (2006). Population structure and eigenanalysis. *Plos Genetics* .
- Papini, R., Renzoni, G., Malloggi, M., & Casarosa, L. (1995). Visceral larva migrans in mice experimentally infected with Baylisascaris transfuga (Ascarididae: Nematoda). *Parassitologia*, *36* (3), 321-329.
- Park, D., O'Doherty, I., Somvanshi, R. K., Bethke, A., Schroeder, F. C., Kumar, U., & Riddle, D. L. (2012). Interaction of structure-specific and promiscuous G-protein-coupled receptors mediates small-molecule signaling in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, *109* (25), 9917-9922.
- 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.

- Rawlings, N. D., & Morton, F. R. (2008). The MEROPS batch BLAST: A tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie*, 90 (2), p.243-259.
- S, W. D., D, F. Y., C, G. A., J, L. E., Ana, M., R, G. J., . . . Zinat, S. (2017). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* (D1), D1.
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., . . . Mcdonald, G. J. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419 (6909), 832-837.
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., . . . Gaudet, R. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449 (7164), 913-918.
- Sameer, V., Glen, V. G., Younes, A., Battle, G. M., Berrisford, J. M., Conroy, M. J., . . . Pauline, H. (2016). PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Research* (D1), D385-D395.
- Schaul, J. (2006). *Baylisascaris transfuga in captive and free-ranging populations of bears (Family: Ursidae)*. The Ohio State University,
- Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46 (8), 919-925.
- Schiffer, P. H., Kroiher, M., Kraus, C., Koutsovoulos, G. D., & Schierenberg, E. (2013). The genome of Romanomermis culicivorax: Revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *Bmc Genomics*, 14 (1), 923.
- Schumacher, T., & Benndorf, R. A. (2017). ABC transport proteins in cardiovascular disease—A brief summary. *Molecules*, 22 (4), 589.
- serine-threonine kinase KIN-29 modulates TGFbeta signaling and regulates body size formation.
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., . . . Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology*, 16 (1), 259.
- Simao, F. A., Waterhouse, R. M., Panagiotis, I., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* (19), 19.
- Sims, J. A., Parsons, J. L., Bissell, H. A., Sikes, R. S., Ouellette, J. R., & Rude, B. J. (2007). Determination of bamboo-diet digestibility and fecal output by giant pandas. *Ursus*, 18 (1), 38-45.
- Smit, A., Hubley, R., & Green, P. (2015). RepeatModeler Open-1.0. 2008–2015. *Seattle, USA: Institute for Systems Biology*. Available from: <httpwww.repeatmasker.org>, Last Accessed May, 1 , 2018.
- Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information & Modeling* , 2324.
- Suhong, X., & Chisholm, A. D. (2012). A Gαq-Ca signaling pathway promotes actin-mediated epidermal wound closure in C. elegans. *Current Biology*, 21 (2), 1960-1967.
- Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current protocols in bioinformatics / editorial board, Andreas D. Baxeavanis . . . [et al.]*, Chapter 4 (Unit 4), Unit 4.10.
- Tempel, S. (2012). Using and understanding RepeatMasker. In *Mobile Genetic Elements* (pp. 29-51): Springer.

- Trapnell, C., Pachter, L., & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25 (9), 1105-1111.
- Tyagi, R., Seshadri, S., Parkinson, J., & Mitreva, M. (2018). Comparative analysis of metabolism in parasitic worms.
- Wang, C., Lan, J., Shen, F., Li, L., Huang, W., Zhi, Y., . . . Li, M. (2015). Toxic Shock of Giant Pandas Caused by Ivermectin. *Chinese Journal of Wildlife* .
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., . . . Guo, H. (2012). MCSscanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40 (7), e49-e49.
- Wildt, D. E., Zhang, A., Zhang, H., Janssen, D. L., & Ellis, S. (2006). *Giant Pandas: biology, veterinary medicine and management* : Cambridge University Press.
- Williamson, A. L., Lustigman, S., Oksov, Y., Deumic, V., Plieskatt, J., Mendez, S., . . . Loukas, A. Ancylostoma caninum MTP-1, an Astacin-Like Metalloprotease Secreted by Infective Hookworm Larvae, Is Involved in Tissue Migration.
- Wise, M. E., Sorvillo, F. J., Shafir, S. C., Ash, L. R., & Berlin, O. G. (2005). Severe and fatal central nervous system disease in humans caused by Baylisascaris procyonis, the common roundworm of raccoons: a review of current literature. *Microbes and Infection*, 7 (2), 317-323.
- Wolf, K. N., Lock, B., Carpenter, J. W., & Garner, M. M. (2007). Baylisascaris procyonis Infection in a Moluccan Cockatoo (Cacatua moluccensis). *Journal of Avian Medicine & Surgery*, 21 (3), 220-225.
- Wu, Y. C., & Horvitz, H. R. (1998). The C. elegans cell corpse engulfment gene ced-7 encodes a protein similar to ABC transporters.
- Xie, Y., Zhang, Z., Wang, C., Lan, J., Li, Y., Chen, Z., . . . Gu, X. Complete mitochondrial genomes of Baylisascaris schroederi, Baylisascaris ailuri and Baylisascaris transfuga from giant panda, red panda and polar bear. *482* (1-2), 0-67.
- Xie, Y., Zhang, Z., Wang, C., Lan, J., Li, Y., Chen, Z., . . . Gu, X. (2011). Complete mitochondrial genomes of Baylisascaris schroederi, Baylisascaris ailuri and Baylisascaris transfuga from giant panda, red panda and polar bear. *Gene*, 482 (1-2), 0-67.
- Xu, M., Molento, M., Blackhall, W., Ribeiro, P., Beech, R., & Prichard, R. (1998). Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog. *Molecular and biochemical parasitology*, 91 (2), 327-335.
- Yang, G. (1995). The Morphology of Endocrine Cells in the Gut Mucosa of the Giant Panda. *Actacentiarum Naturalum Universitatis Pekinesis* .
- Yang, G. (1998). Advances on parasites and parasitology of Ailuropoda melanoleuca. *Chin J Vet Sci*, 18 , 206-208.
- Zhang, J.-S., Daszak, P., Huang, H.-L., Yang, G.-Y., Kilpatrick, A. M., & Zhang, S. (2008). Parasite threat to panda conservation. *Ecohealth*, 5 (1), 6-9.
- Zhang, L., Yang, X., Wu, H., Gu, X., Hu, Y., & Wei, F. (2011). The parasites of giant pandas: individual-based measurement in wild animals. *Journal of Wildlife Diseases*, 47 (1), 164-171.
- Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., . . . Fan, W. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, 45 (1), 67-U99.
- Zhao, X., & Hao, W. (2007). LTR.FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* (suppl.2), suppl.2.

Zhou, C., Hu, J., Yuan, C., & Wei, F. (1997). Giant pandas food habits and feeding behaviour in Mabian Dafengding Natural Reserve. *Journal of Sichuan Teachers College* .

Zhu, X. Q., Korhonen, P. K., Cai, H., Young, N. D., Nejsun, P., Von Samson-Himmelstjerna, G., . . . Min, J. (2015). Genetic blueprint of the zoonotic pathogen *Toxocara canis*.

Zou, Xinghuai, Wang, Aimin, Zeng, Lujin, . . . Kongju. (1998). Lethal factors of diseases and protective countermeasures of wild and penned Giant pandas. *Journal of Forestry Research* .

Zumla, A., & Hui, D. S. (2019). Emerging and reemerging infectious diseases: global overview. *Infectious Disease Clinics*, 33 (4), xiii-xix.

## Tables

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

Description	<i>B. schroederi</i> (this study)	<i>B. schroederi</i> <sup>#</sup>	<i>A.suum</i>	<i>P. u</i>
Genome size (bp)	262,743,947	281,639,769	265,545,801	253,
Number of Scaffolds; Contigs	106; 639	2,778; 15,567 (>1000 bp)	31,538; 40,509	1,27
Average length of Scaffolds; Contigs (bp)	2,478,717; 410,763	-	8,420; 6,506	198,
Gap length (bp; % of genome)	266,500 (0.10%)	13,257,555 (4.70%)	1,980,846 (0.7%)	966,
N50 of Scaffolds; Contigs (bp)	12,688,039; 1,154,808	888,870; 42,126	290,558; 46,632	1,82
N90 of Scaffolds; Contigs (bp)	7,262,801; 161,963	104,281; 7439	48,674; 10,466	204,
Genome GC content (%)	37.34	37.26	37.85	39.0
Repetitive sequences (%)	9.53	-	4.4	-

<sup>#</sup> The published genome information of *B. schroederi* (not released)(Y. Hu et al., 2020).

## FIGURE LEGENDS

**Fig.1 The phylogenetic relationships among six nematodes and genomic characteristics and synteny of *B. schroederi*.**(a) Genomic characterization of *B. schroederi* genome. The figure shows the gene number, repeat content, GC content, sequencing coverage and scaffolds from the center to the edge. (b) Synteny of *B. schroederi* with *P. univalens* and *T. canis* at the gene level. Different colors represent different synteny blocks.(c) Upset plot showing the intersection of gene family expansions in nematodes. Each row represents a nematode. Black circles and vertical lines between the rows represent the intersection of expanded families between species. The barplot indicates the total gene family count in each intersection. (d) Time-calibrated maximum likelihood phylogenetic tree of six nematodes. The estimated divergent times are shown at the bifurcations. The numbers below the nodes represent the number of gene families significantly expanded, maintained, and contracted, respectively.

**Fig.2 The expansion and contraction of roundworm gene families.**(a) Significant increases and decreases in roundworm gene families. The solid circle and the solid triangle represent the top KEGG pathways that are enriched in the expanded gene families of Ascariasis compared with *C. elegans* or *M. hapla* , respectively. The open circle and the open triangle represent the top KEGG pathways that are enriched in the contracted gene families of Ascariasis compared with *C. elegans* or *M. hapla* , respectively. (b) GO function enrichment and gene copy number of the significantly expanded gene families in roundworms; (c) The proportion of GO functional genes in the gene family with significant expansion (or contraction) in roundworms compared to the total number of expansion (or contraction) genes. The red asterisk represents the p value of statistical Sidak's multiple comparisons tests of expansion and contraction of genes comparing with *C. elegans* or *M. hapla*(One asterisk represent  $10^{-1}$ ).

**Fig.3 Expansion and contraction of *B. schroederi* gene families compared with three roundworms (*A. suum* *P. univalens* and *T. canis*).** (a) Enrichment of the KEGG pathway in some significantly



expanded gene families of *B. schroederi*. The proportion represents the ratio of the number of expanded genes located in the pathway (target genes) to all genes in the pathway (background genes). (b) REVIGO clusters of significantly over-represented GO items for significantly expanded gene families in *B. schroederi*. The position of the bubbles is based on semantic similarity of GO terms. (c) Enrichment of KEGG pathways in *B. schroederi*'s unique gene families. (d) Heatmap showing the gene families of *B. schroederi* that are significantly expanded or contracted. The x axis represents the four roundworms of Ascariasis, whereas the y axis represents the families.

**Fig.4 Life history of *B. schroederi* and the effect of actin gene on muscle contraction.** (a) Life history of *B. schroederi*. L1 and L2 represent in vitro developmental stages, and L2 larvae enter the host body after developing into the infective stage. L3 and L4 represent the stage of internal organ migration of the larva. Stage L5 larvae return to the small intestine and develop into adult worms through sexual maturation. (b) Schematic diagram of anatomical cross-section of *B. schroederi*; (c) Multiple signaling pathways are involved in actin polymerization, and genes in red are positively selected genes (PSGs). It shows significant expansion of three key gene families involved in actin polymerization.

**Fig.5 Demographic history of the *B. schroederi* reconstructed from the reference and population resequencing genomes.** (a) The red and purple line represent the estimated effective population size of *B. schroederi* and host, respectively. The 100 gray curves of *B. schroederi* and host represent the PSMC estimates for 100 sequences randomly resampled from the original sequence. Generation time (g) of *e* and giant panda were 0.17 and 12 years, respectively. The neutral mutation rate per generation ( $\mu$ ) of *B. schroederi* and giant panda were  $0.9 \times 10^{-8}$  and  $1.3 \times 10^{-8}$ , respectively. The black line shows the MAR of Chinese loess. (b) Longitudinal change of the effective population size of the *B. schroederi* populations. The effective population sizes ( $N_e$ ) were estimated using the MSMC2 method. QLI: Qinling population; SC: Sichuan population.

**Fig. 6 Population structure and relationships of Sichuan (SC) in comparison to Qinling (QLI) population.** (a) The geographic distribution of the sampling locations for QLI and SC populations. (b) PCA analysis of two populations; (c) A maximum likelihood (ML) phylogenetic tree with 100 bootstrap tests constructed using whole-genome SNPs information. We used *P. univale* as the outgroup; (d) Population structure of SC and QLI populations (K from 2 to 5). The y axis quantifies the proportion of the individual's genome from inferred ancestral population, and x axis shows the different individuals.

**Fig.7 Analysis of natural selection in captive populations.** (a) Genomic regions with selection sweep signals in captive (SC) and wild (QLI) *B. schroederi* population. Distribution of  $\ln$  ratio ( $\partial\pi, \omega_{\text{AID}}(\text{XAI})/\partial\pi, \omega_{\text{AID}}(\Sigma^*)$ ) and  $F_{\text{st}}$  of 50 kb windows with 10 kb steps. Red dots represent windows fulfilling the selected regions requirement (corresponding to Z test  $P < 0.005$ , where  $F_{\text{st}} > 0.21$  and  $\ln$  ratio  $> 0.34$ ). (b) Plot of iHS showing loci under positive selection of captive (SC) population. SNPs with  $|iHS| \geq |iHS_m|$  (3.89, top 1%) were shown above the dashed horizontal line. Nucleotide diversity around *glc-1*, *nrf-6*, ABC transporter *ced7* and *bre-4* loci using 10-kb sliding windows were displayed above the genes. The decay of haplotype homozygosity around a focal marker were displayed on the right side of the figure. The furcation structures represent the complete information contained in the concept of extended shared haplotypes EHH (Pardis C. Sabeti et al., 2002). The root (focal marker) is indicated by a vertical dashed line. The thickness of the lines corresponds to the number of scaffolds sharing a haplotype. (c) XP-EHH from each SNP core showing the same nucleotide between the subject and the comparison target, also transformed to  $P$  values and plotted in logarithmic scale.

**Fig.8 The position of known and potential drug target genes on superscaffolds.** Different colors indicate different known drugs, and black indicates potential drug targets. The chemical structural formulas of 23 known drugs are drawn. The circles following the potential drug targets represent the six criteria, with a red solid circle indicating match condition and a hollow circle indicating mismatch condition. the six criteria were: (1) Similarity with ChEMBL known drug targets having a highly conserved alignment ( $> 80\%$ ); (2) Lack of human homologues; (3) Related to lethal, L3 arrest, flaccid, molt defect or sterile phenotype; (4) A predicted metabolic chokepoint; (5) A predicted excretory/secretory protein (EP); (6) The

protein has a structure in the PDB. Potential drug target proteins encoding genes on each superscaffold and corresponding scores are marked (black).









