Received: 15 August 2022 | Revised: 6 January 2023 | Accepted: 30 January 2023

DOI: 10.1111/1755-0998.13760

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Genome-wide exploration reveals distinctive northern and southern variants of *Clonorchis sinensis* in the Far East

Liina Kinkar[1,2] | Pasi K. Korhonen[1] | Urmas Saarma[2] | Tao Wang[1] | Xing-Quan Zhu[3] | Ivon Harliwong[4] | Bicheng Yang[4] | J. Lynn Fink[4,5] | Daxi Wang[1,6] | Bill C. H. Chang[1,7] | Galina N. Chelomina[8] | Anson V. Koehler[1] | Neil D. Young[1] | Robin B. Gasser[1]

[1]Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, Australia

[2]Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia

[3]College of Veterinary Medicine, Shanxi Agricultural University, Taigu, China

[4]BGI Australia, Herston, Queensland, Australia

[5]The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Woolloongabba, Queensland, Australia

[6]BGI-Shenzhen, Shenzhen, China

[7]Yourgene Health, Taipei, Taiwan

[8]Department of Parasitology, Federal Scientific Center of the East Asia Terrestrial Biodiversity FEB RAS, Vladivostok, Russia

Correspondence
Neil D. Young and Robin B. Gasser, Department of Veterinary Biosciences, Melbourne Veterinary School, The University of Melbourne, Parkville, VIC, Australia.
Email: nyoung@unimelb.edu.au and robinbg@unimelb.edu.au

Funding information
Australian Research Council, Grant/Award Number: LP180101085,LP180101334, LP190101209; The Estonian Ministry of Education and Research, Grant/Award Number: PRG1209; The Fund for Shanxi "1331 Project", Grant/Award Number: 20211331-13; The Special Research Fund of Shanxi Agricultural University for High-level Talents, Grant/Award Number: 2021XG001

Handling Editor: Sebastien Calvignac-Spencer

Abstract

Clonorchis sinensis is a carcinogenic liver fluke that causes clonorchiasis—a neglected tropical disease (NTD) affecting ~35 million people worldwide. No vaccine is available, and chemotherapy relies on one anthelmintic, praziquantel. This parasite has a complex life history and is known to infect a range of species of intermediate (freshwater snails and fish) and definitive (piscivorous) hosts. Despite this biological complexity and the impact of this biocarcinogenic pathogen, there has been no previous study of molecular variation in this parasite on a genome-wide scale. Here, we conducted the first extensive nuclear genomic exploration of *C. sinensis* individuals (n = 152) representing five distinct populations from mainland China, and one from Far East Russia, and revealed marked genetic variation within this species between "northern" and "southern" geographical regions. The discovery of this variation indicates the existence of biologically distinct variants within *C. sinensis*, which may have distinct epidemiology, pathogenicity and/or chemotherapic responsiveness. The detection of high heterozygosity within *C. sinensis* specimens suggests that this parasite has developed mechanisms to readily adapt to changing environments and/or host species during its life history/evolution. From an applied perspective, the identification of invariable genes could assist in finding new intervention targets in this parasite, given the major clinical relevance of clonorchiasis. From a technical perspective, the

genomic-informatic workflow established herein will be readily applicable to a wide range of other parasites that cause NTDs.

KEYWORDS
carcinogenic liver fluke, clonorchiasis, *Clonorchis sinensis*, Far East Asia, genetic differentiation, population genomic variation

1 | INTRODUCTION

Neglected tropical diseases (NTDs) represent a group of 20 conditions that are common predominantly in tropical and subtropical regions of the world, mostly affecting people in disadvantaged communities, with major health and socioeconomic consequences for >1 billion people (WHO, 2020). Of particular concern are parasitic flatworms (trematodes), such as the blood fluke—*Schistosoma haematobium*—and liver flukes—*Opisthorchis viverrini* and *Clonorchis sinensis*—which, in addition to causing chronic urogenital or hepatic diseases, induce malignant cancers in people (IARC, 2012).

In particular, *C. sinensis*, colloquially called the Chinese liver fluke, is a food-borne parasite causing clonorchiasis in ~35 million people worldwide, predominantly in parts of China (southeastern and northeastern provinces), Vietnam, Korea and the Russian Far East (southern region) (Lun et al., 2005). Humans and other piscivorous mammals (e.g., felids or canids) acquire the infection by eating raw or undercooked freshwater (predominantly cyprinid) fish containing encysted larvae (metacercariae), which excyst and develop into adult worms in the biliary system of the liver and/or pancreatic ducts (Lun et al., 2005). Chronic infection with *C. sinensis* often leads to hepatobiliary complications, such as chronic inflammation, fibrosis and cholelithiasis, and can induce malignant bile-duct cancer (cholangiocarcinoma) (Kim et al., 2016; Qian et al., 2016; Tang et al., 2016). Despite an effective drug—praziquantel—being available to treat *C. sinensis* infection and reduce transmission (Yangco et al., 1987), persistent cultural habits, such as the deeply rooted social custom of eating raw fish, impede efforts to control this liver fluke in endemic regions (Hong et al., 1998; Qian, 2014; Qian et al., 2016).

Although some aspects of the epidemiology of clonorchiasis are understood (Lun et al., 2005; Qian et al., 2012), very little is known about the genetics of *C. sinensis*, including the genetic make-up of populations, and whether population variants exist that might differ in their biology, transmission, virulence/pathogenicity and/or susceptibility to anthelmintics (e.g., praziquantel). Extending early studies of *C. sinensis*, conducted mainly using relatively small numbers of genetic loci and/or limited sample sizes (reviewed by Wang, Young, et al., 2018), we discovered unexpectedly high levels of mitochondrial (mt) genetic variability within *C. sinensis* (see Kinkar et al., 2020). However, given the likely unique mutation rates and mode of inheritance in mtDNA (Boore, 1999; Moore, 1995), it has been unclear whether the nature and extent of this substructuring is reflected in nuclear DNA (cf. Tatonova et al., 2012). Building on recent genomic work over the past decade (Huang, Chen, et al., 2013; Wang et al., 2011; Wang, Korhonen, et al., 2018; Yoo et al., 2011; Young et al., 2010, 2021), we investigated here, on a genome-wide scale, the nature and extent of nuclear genetic variation within *C. sinensis* from endemic regions in China and the Russian Far East, using an advanced bioinformatic workflow. We discuss the findings in relation to the biology of *C. sinensis*, the epidemiology and pathogenesis of clonorchiasis, and the implications that this variation might have for the control of this parasite.

2 | MATERIALS AND METHODS

2.1 | Sequencing of genomic DNA samples

Genomic DNA samples (n = 152) representing adult *Clonorchis sinensis* worms from China (southeastern provinces: Guangdong, Guangxi and Hunan; northeastern provinces: Heilongjiang and Jilin) and the Russian Far East (Primorsky Krai region) were available from previous studies (Chelomina et al., 2014; Kinkar et al., 2020; Lin et al., 2011) (Table S1). Total genomic DNA had been extracted from the posterior tip (2–4 mm) of each of the 152 specimens using the Dneasy Blood & Tissue Kit (Qiagen), according to the manufacturer's protocol; individual DNA samples were whole-genome amplified (using the REPLI-g Mini Kit, Qiagen; cat. No. 150025) to produce individual genomic DNA libraries and then sequenced (100-bp paired-end reads) using the BGISEQ-500 platform (BGI Australia) (Kinkar et al., 2020).

2.2 | Recording nucleotide variation

A newly-established pipeline, called *Escalibur* (Korhonen et al., 2022), was used to record single nucleotide polymorphisms (SNPs) and insertion/deletion events (indels) at individual positions in relation to a reference nuclear genome of *C. sinensis* from China (C_sinensis-2.0; ~547 Mb in size; GenBank accession no. GCA_000236345.1; Huang, Chen, et al., 2013). Raw DNA sequence data in FASTQ format (Cock et al., 2010) for each of the 152 *C. sinensis* specimens were filtered for quality (Phred quality score cut-off: 20) and trimmed using the program TRIMMOMATIC version 0.38 (Bolger et al., 2014). Filtered read-pairs were mapped to the reference genome sequence using the Burrows-Wheeler Aligner (BWA) version 0.7.8 (Li & Durbin, 2009). The mapped data were then used to record SNPs and indels using the Genome Analysis Toolkit (GATK version 4.0.8.1; McKenna et al., 2010). "Raw" SNPs were then selected and filtered for quality using GATK VARIANTFILTRATION and VCFTOOLS version 0.1.16 (Danecek et al., 2011); SNPs were

retained if the following criteria were met: variant confidence (QD) >2.0; strand bias (FS) <60.0; mapping quality (MQ) >40.0; mapping quality (MQRankSum) >−12.5; read position bias (ReadPosRankSum) >−8.0; variant quality (minQ) >30; and genotype quality (minGQ) >30. All SNPs were contextualized (including exons, introns and intergenic elements) using SNPEFF version 5.0 (Cingolani et al., 2012) employing a GFF annotation file available for the reference genome.

For the selection of protein-coding genes for analyses of nucleotide variation, read coverage and depth across the coding regions were assessed using MOSDEPTH version 0.3.1 (Pedersen & Quinlan, 2018), considering reads with a mapping quality (MQ) of >30. Genes were selected if read coverage was >95% (minimum depth: 5) and average coverage was ≥20 across the coding domain for all samples sequenced. Subsequently, for each individual sample, filtered SNPs were transferred to the genome reference using GATK FASTAALTERNATEREFERENCEMAKER version 4.1.0.0; heterozygous positions were given IUPAC codes (Cornish-Bowden, 1985). The predicted sequence of the coding domain of each gene was extracted using BEDTOOLS version 2.29.0 (Quinlan & Hall, 2010).

2.3 | Identifying variable and invariable protein-coding genes

To investigate sequence variation and conservation in protein-coding genes of *C. sinensis*, nucleotide diversity values—the average proportion of distinct nucleotides between all pairs of sequences obtained for a population (Nei, 1987)—were calculated for each protein gene. Genes were ranked from least to most variable, according to nucleotide diversity and the standard deviation thereof; genes in the bottom and top 10% were designated as "invariable" and "variable", respectively. Amino acid sequences inferred from these genes were classified based on their homology (BLASTP; *E*-value $<10^{-8}$) to sequences in the KEGG database. Enriched protein families and/or biological pathways were defined for gene sets using Fisher's exact test, employing a custom script and linking data to KEGG biological pathways and the KEGG BRITE hierarchy.

2.4 | Analyses of nuclear genomic protein-coding gene sets

To estimate the genetic relationships among *C. sinensis* individuals, BCFTOOLS version 1.14 (Danecek et al., 2021) was used to remove SNPs in linkage disequilibrium (LD; r^2 cutoff .2). Subsequently, unlinked variants were subjected to phylogenetic analyses using Bayesian inference (BI) in BEAST2 (Bouckaert et al., 2014), employing the general time-reversible model (GTR; Tavaré, 1986) of nucleotide substitution. The number of Markov chain Monte Carlo (MCMC) iterations (Geyer, 1991; Hastings, 1970; Metropolis et al., 1953) was 100,000,000, from which the first 80% were discarded as nonconvergent "burn-in." Five independent runs were combined using LOGCOMBINER version 2.6.6, and the resultant tree was produced using TREEANNOTATOR version 2.6.6 (Drummond & Rambaut, 2007). Nodes with a posterior probability (pp) of <0.8 were collapsed using TREEGRAPH 2 version 2.15.0 (Stöver & Müller, 2010), and the final consensus tree was drawn in FIGTREE version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Subsequently, the R (R Core Team, 2018) packages VCFR version 1.12.0 (Knaus & Grünwald, 2017) and ADEGENET version 2.1.5 (Jombart, 2008) were used to infer population genetic differentiation by principal component analysis (PCA) and discriminant analysis of principal components (DAPC; Jombart et al., 2010). For PCA, the numbers of principal components (PCs) retained were determined based on eigenvalues and percentages of variance for individual PCs. Specifically, PCA results were plotted, retaining: (i) the first two PCs accounting for the largest percentage of variance, (ii) all PCs retaining >1% of variance, (iii) all PCs to the left of the "elbow" point in the decreasing curve of variance and (iv) all PCs retaining up to 30%, 50% or 70% of total variance (Jolliffe, 1986). Dimensionality of more than two PCs was reduced for two-dimensional plotting using Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018), employing the R package UMAP v.0.2.7.0. DAPC analyses were conducted twice. In the first run, the sampling locations pertaining to the Chinese provinces of Guangdong, Guangxi, Hunan, Jilin and Heilongjiang, and the Primorsky Krai region in the Russian Far East were used as a priori groups. In the second run, the "find.clusters" function in the ADEGENET package was used to determine the number of groups de novo, with an optimal number selected based on the lowest value of the Bayesian Information Criterion (BIC). For both analyses, the optimal number of PCs was determined using a cross-validation test with the "xvalDapc" function in ADEGENET; the number of replicates at each level of PC retention was 100. Plots were generated using the R package GGPLOT2 (Wickham, 2016).

2.5 | Analysis of PC-correlated SNPs in PCA

To investigate the SNPs linked to population differentiation along PCs, first, the function "snpgdsPCA" available in the R package SNPRELATE version 1.28.0 (Zheng et al., 2012) was used to calculate eigenvectors using all SNP positions identified. Then, the function "snpgdsPCACorr" was used to investigate the correlation between SNPs and the eigenvector of the first PC. All SNPs with a correlation value of ≥.5 were selected; the protein genes in which these SNPs occurred were linked to KEGG biological pathways and the KEGG BRITE hierarchy (see section "Identifying variable and invariable protein-coding genes").

3 | RESULTS

3.1 | Sequence data sets and mapping results

The 152 high-quality, short-read genomic sequence data sets (37.2–110.4 Gb; median: 67.8 Gb) representing the 152 specimens of *Clonorchis sinensis* were generated and individually mapped to the reference nuclear genome. This effort yielded 325.7–939.7 million

(median: 636.1) aligned sequence reads for individual samples, with 82.8%–98.5% of reads mapping as pairs, with 58- to 170-fold coverage of the reference genome sequence (Table S2). These data were then used to estimate the nature and extent of nuclear genetic variation among the 152 individuals of *C. sinensis*.

3.2 | SNP numbers in protein-coding genes

In total, 3,976,687–6,754,432 high-quality, filtered SNPs were recorded for the 152 individual *C. sinensis* specimens in relation to the reference genome sequence (Table S3). Of these SNPs, 2,320,714–3,941,970 (57.5%–59.0%), 1,565,181–2,658,718 (38.7%–40.2%) and 90,792–153,744 (2.2%–2.3%) per individual were within intergenic, intronic and exonic regions, respectively. Based on read coverage and depth, a set of 9409 protein-coding genes was retained for analyses. In total, 675,728 SNP positions were detected within all coding domains, 673,905 (99.7%) and 125,466 (18.6%) of which were heterozygous and homozygous, respectively, in one or more individuals. On an individual specimen level, *C. sinensis* sample no. 188 from a dog from Hunan had the lowest percentage (60.6%) of heterozygous SNPs (n = 37,624), whereas sample no. 121 from a dog from Guangdong had the highest percentage (93.2%; n = 96,643) (Table S3). The average read depth at the 675,728 SNP positions in *C. sinensis* individuals ranged from 40 to 140; the percentage of quality-filtered SNPs (i.e., those removed from analysis) per individual ranged from 0.03% to 3.23% (Table S4).

3.3 | Invariable and variable genes, and their association with biological pathways in *C. sinensis*

In total, 941 and 941 "variable" and "invariable" protein-coding genes were identified to represent all 152 *C. sinensis*individuals studied. Analyses of respective protein families and biological pathways revealed that the 941 invariable protein-coding genes were associated with functions including genetic information processing (e.g., the spliceosome protein complex, mitochondrial biogenesis, ubiquitin-mediated proteolysis and mRNA surveillance), signalling and cellular processes (e.g., GTP-binding proteins, pore-forming ion channels and mitophagy), and cellular respiration and metabolism (e.g., glucagon signalling, including phosphoenolpyruvate carboxykinase—gene csin103490; Huang, Chen, et al., 2013) (Figure 1; Tables S5 and S6). Among the invariable genes was also one predicted to code for a myophilin-like protein (csin111322; Huang, Chen, et al., 2013). On the other hand, the 941 variable protein-coding genes were associated with functions including signalling and cellular processes (e.g., the Niemann-Pick type C2 [NPC2]-like proteins), cellular transport and catabolism (e.g., autophagy and lysosomal pathways) as well as genetic information processing (e.g., transcription factors) (Figure 1; Tables S7 and S8). The most variable genes (i.e., those to the right of the "elbow" point in the increasing curve of nucleotide diversity values; n = 250; Figure 1) are listed in Table S9.

3.4 | Clear genetic differentiation between northern and southern variants of *C. sinensis*

In total, 228,275 unlinked SNPs were used to assess the genetic relationships among all 152 *C. sinensis* individuals included in the study. Phylogenetic relationships inferred using the Bayesian approach revealed three main clusters of *C. sinensis* individuals with absolute nodal support (pp = 1.0) (Figure 2; Figures S1–S5). These *C. sinensis* clusters were linked to geographical origin: clade 1 from the provinces Guangdong and Guangxi in the south of China; clade 2 from Hunan also in the south; and clade 3 from Jilin and Heilingjiang in the north of China as well the adjacent Primorsky Krai region of the Russian Far East. Within clade 3, all samples from the latter region clustered together (pp = 1.0) to the exclusion of those from Heilongjiang and Jilin. The phylogenetic relationships were supported by PCA. As the percentage of total variance of individual PCs was low (<3%; Figure S6), different numbers of PCs were retained and plotted to assess the consistency of the results (Figure 2; Figures S7–S9). These sets consistently represented the three distinct clades in the phylogenetic tree. Sets of 11, 61 and 92 PCs, associated with 14%, 50% and 70% of the total variance in the data, respectively, revealed a minor distinction between individuals from the Primorsky Krai region and other individuals within clade 3 (Figures S8 and S9).

The DAPC analysis using geographically predefined clusters also revealed a distinction among clades 1, 2 and 3, and a uniqueness of *C. sinensis* individuals originating from Primorsky Krai (Figure 2; Figure S11). For this analysis, a set of 61 PCs relating to 50% of the variance was determined as the optimum number of components, as inferred from a DAPC cross-validation test (Figure S10). Without predefined groups, the optimum number of groups was determined as two (Figure 2; see Figure S12 for a BIC plot), pertaining to samples obtained from southern provinces (Guangdong, Guangxi and Hunan), and from northern provinces (Heilongjiang and Jilin) of China and the Primorsky Krai region in Russia. The optimum number of PCs with de novo group designations ranged between one and 64 (see Figure S13 for cross-validation results). The analysis was run using 61 PCs—as for a priori groupings—confirming the genetic distinctiveness of southern and northern *C. sinensis* along a single discriminant function (Figure 2; Figure S14).

3.5 | Higher heterozygosity in a set of SNPs in northern *C. sinensis*

The apparent genetic distinctiveness of southern and northern *C. sinensis* was assessed further by correlating variation in protein genes with the PC differentiating the two populations. Results of

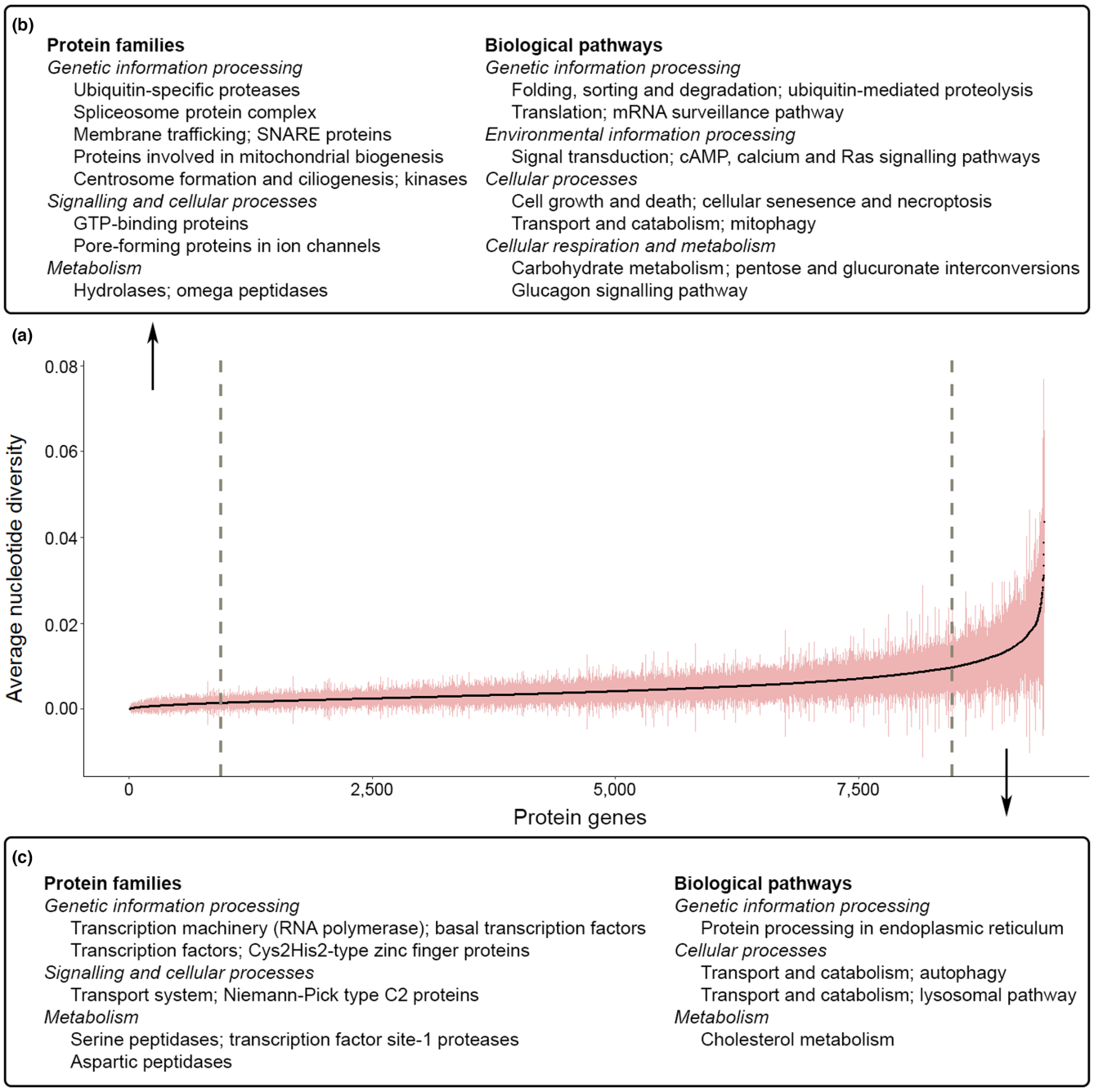

FIGURE 1 Sequence conservation and variation in 9409 protein-coding genes in *Clonorchis sinensis*. (a) Ranked protein genes, according to nucleotide diversity values (in black) across coding domains. Pink lines represent standard deviation values; dashed lines demarcate genes designated as "invariable" (conserved) and "variable." A summary of significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) biological pathways and protein families encoded by (b) invariable or (c) variable genes are listed. Cyclic adenosine monophosphate (cAMP); guanosine triphosphate (GTP).

PCA using all identified SNP positions (n = 675,728) were in accordance with those calculated from unlinked SNPs (n = 228,275; cf. Figure 2), revealing a clear genetic distinction between the northern and southern samples along PC1 (Figure S15). A total of 1623 SNPs correlated highly for PC1 (correlation coefficient of ≥.5; Figure S16). A pairwise comparison of these SNPs among *C. sinensis* individuals revealed more heterozygosity for the northern Chinese provinces of Heilongjiang and Jilin, and the Primorsky Krai in Russia (average number of heterozygous loci per specimen = 752) than for the southern Chinese provinces Guangdong, Guangxi and Hunan (average number of heterozygous loci per specimen = 269) (Figure 3). These 1623 SNPs were in 750 protein-coding genes, inferred to be associated with functions including genetic information processing (e.g., spliceosome-associated proteins), cellular signalling and transport (e.g., electrochemical potential-driven transporters, monocarboxylate transporters and rhodopsin G protein-coupled receptors),

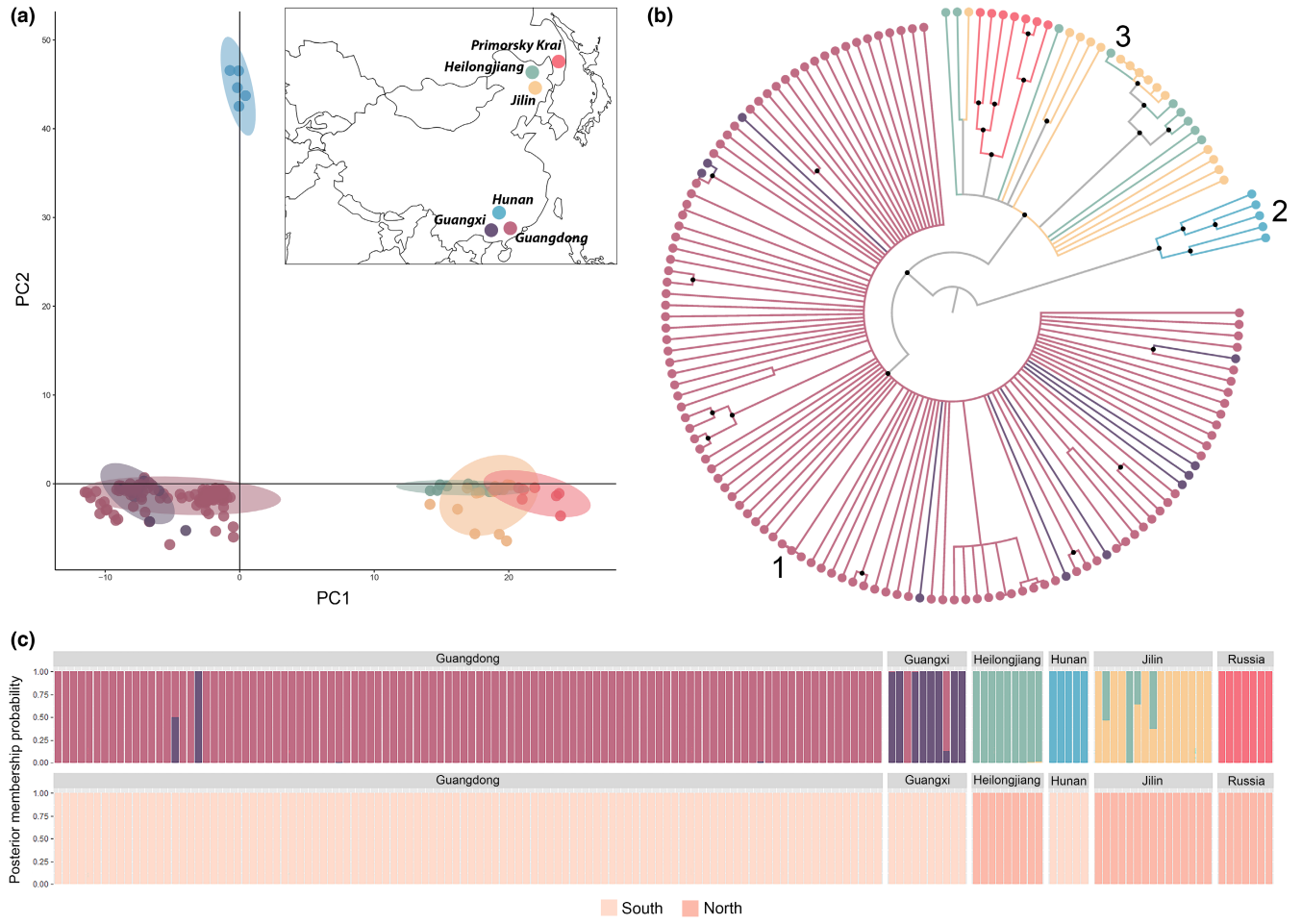

FIGURE 2 Population structure of *Clonorchis sinensis*. (a) Principal component analysis (PCA) of genetic differentiation using the first two principal components (PCs) accounting for the largest percentage of variance (Figure S6). Points represent individual samples; colours pertain to the geographical locations from which they originate (a map of sampling localities are shown at the top right). Ellipses enclose 95% of specimens of *C. sinensis* originating from a particular geographical location. (b) A consensus phylogenetic tree inferred by Bayesian inference analysis. Branches are coloured based on the geographical origin of specimens. Small black dots at the nodes indicate absolute posterior probability values (pp = 1.0). Nodes with a pp of <0.8 were collapsed. Clades 1–3 are indicated. (c) Barplots of posterior membership probabilities inferred by the discriminant analysis of principal components (DAPC) method. Each individual is represented by a vertical line, coloured in accordance with the genetic cluster to which it was assigned. The result of the analysis using sampling locations as a priori groups is shown at the top; the bottom plot represents de novo group assignments ("south" and "north"; respective colours are indicated at the bottom).

cellular catabolism (e.g., mitophagy and lysosomal proteins), and protein metabolism and processing (e.g., cystein peptidases and serine/threonine phosphatases) (Tables S10 and S11).

4 | DISCUSSION

The present assessment of genetic variation within *Clonorchis sinensis* using 152 individual specimens from endemic regions revealed a clear genetic differentiation between northern and southern populations, with markedly higher levels of heterozygosity (allelic variability) in SNPs linked to 750 protein genes encoded in the nuclear genome of northern compared with southern "worms" (Figures 2 and 3), contrasting a lack of differentiation inferred previously using mitochondrial genomic data (Kinkar et al., 2020).

We hypothesize that this distinctive pattern of nuclear genetic variability relates to the parasite's respective adaptation to these two disparate geographical regions, characterized by vastly different climates, which likely have an impact on the types of intermediate host species (snails and fish) involved in the life cycle and transmission, and, more broadly, might effect the epidemiology and ecology of the parasite and disease in these regions (Chelomina, 2018; Wang, Young, et al., 2018). The variation in allelic variability between these populations likely associates with selected biological processes or pathways, inferred here to include cellular transport and signalling (e.g., rhodopsin G protein-coupled receptors), gene splicing, and cellular metabolism and protein degradation (e.g., cysteine proteases) (cf. Tables S10 and S11), which we propose undergo conspicious changes as the parasite adapts to a particular ecological niche. Another hypothesis could be that the genetic differentiation

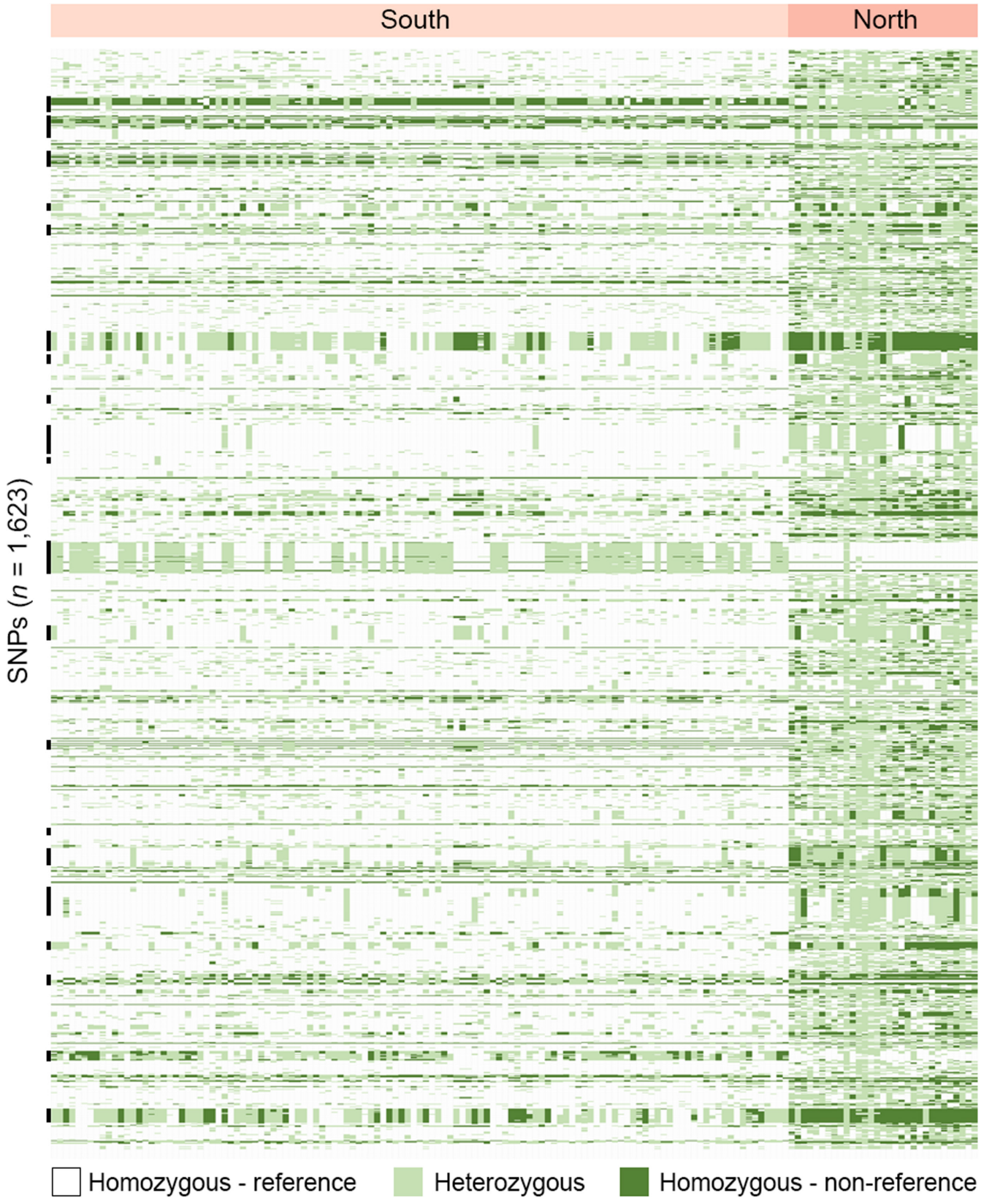

FIGURE 3 Allelic variation for SNPs ($n = 1623$) having the highest correlation with the first principal component calculated using a set of 675,728 SNPs (cf. Figures S15 and S16). Rows indicate individual SNPs and columns indicate individual samples ($n = 152$); their geographical origin—northern or southern population—is indicated at the top. Colours pertain to: (i) nucleotides matching the reference nuclear genome (homozygous–reference; white), (ii) heterozygous positions with one nucleotide matching the reference (light green) and (iii) "fixed" nucleotides that differ from the reference (homozygous–non-reference; dark green). Groups of ≥10 SNPs within 10,000 bp from one another are indicated with a black vertical line on the left.

between northern and southern *C. sinensis* populations is not adaptive, but rather the result of geographical isolation. However, this proposal seems less likely, as infected hosts (including humans) likely disseminate *C. sinensis* when they migrate between or among distinct geographical regions, or via the transportation of freshwater fish or shrimp products as food from one region to another, which would be reflected in limited population genetic differentiation, particularly in the absence of significant local selection pressure.

Interestingly, *C. sinensis* samples from Hunan ($n = 5$; clade 2) were genetically distinct from those representing the other populations studied (Figure 2, Figures S7–S9 and S11), whereas the DAPC analysis assigned samples from Hunan as well as Guangdong and Guangxi ($n = 116$; clade 1) to a "southern" variant of *C. sinensis* (Figure 2). In a previous analysis of mitochondrial data (Kinkar et al., 2020), specimens from Hunan clustered together with *C. sinensis* from South Korea. Thus, it is possible that clade 2 has a broader geographical distribution than Hunan, but this proposal would need to be tested in a future analysis following extended sampling across southern and central China and South Korea.

Despite distinctive patterns of allelic variability in 750 protein-coding genes between southern and northern *C. sinensis*, high levels of heterozygosity (60.6%–93.2% of SNPs in genes; Table S3) were common to all individuals of *C. sinensis* studied here. Inbreeding (either via selfing or fertilization between closely related individuals) is often associated with a loss of heterozygosity in the genome (Charlesworth & Charlesworth, 1987). Although not studied yet in great detail for *C. sinensis*, inbreeding would be expected, given that this digenean is a hermaphrodite and reproduces clonally in the snail host (cf. Wang, Young, et al., 2018). Selfing, presumably "favoured" at low parasite densities in the host, lacks the benefits of genetic mixing via outcrossing, and, in the presence of outcrossing opportunities, the probability of encountering, and mating with, a genetically similar individual in a host seems high (Ramm, 2017). This might occur due to relatively high numbers of clones produced in snails being available to freshwater fish, in turn, giving rise to a relatively genetically homogeneous population of worms in definitive hosts in a particular (endemic) geographical location. The high levels of heterozygosity observed here suggest that a genetic or genomic mechanism exists in *C. sinensis* that is capable of maintaining heterozygosity (cf. Bayne & Grevelding, 2003; Guo et al., 2016) in individual worms in endemic settings. Such heterozygosity has also been described for the related opisthorchid *Opisthorchis felineus*, and hypothesized to relate to the parasite's capacity to rapidly adapt to environmental pressures (Ershov et al., 2019). Clearly, exploring how and why allelic variability is maintained in natural populations of *C. sinensis* would be informative.

The finding of high nucleotide variation in genes encoding NPC2-like proteins (Figure 1; Table S7) agrees with a previous study of *C. sinensis* from China and South Korea (Wang, Korhonen, et al., 2018), and indicates a rapid evolution of this gene family (Anari et al., 2020). NPC2-like proteins are lipid-binding molecules, which have undergone a substantial expansion within *C. sinensis*, and whose functions probably go beyond sterol binding and transport, to enable the parasite to survive in a harsh (nutrient-poor) environment within the biliary system within the host animal (Anari et al., 2020). Among other variable genes were some encoding transcription factors (Figure 1; Table S7), which represent key genomic "switches" orchestrating gene transcription/expression (Maston et al., 2006). Although the impact of mutations in transcription factors on gene regulatory networks is not understood in parasitic worms, and the genetic determinants of the abundances of synthesized proteins in the cell relate to a complex network of interactions (Maston et al., 2006; Williams et al., 2007), the initiation of transcription represents a critical regulatory phase. It would be interesting to explore how transcription factors regulate gene expression and cellular processes in this digenean parasite, and to identify the roles of these factors in adaptation. Given that *C. sinensis* can infect a range of intermediate and definitive (piscivorous) host species (Tang et al., 2016; Wang, Young, et al., 2018), one might expect elevated nucleotide variation in genes governing host invasion and specificity. For instance, in China, at least 102 species of fish and four species of shrimp are known to be second intermediate hosts of *C. sinensis* (see Tang et al., 2016), but nothing is known about the molecular basis of host affiliations or host-parasite interactions. This would be an interesting area to explore in the future.

Conserved genes linked to key biological pathways are likely to be critical for *C. sinensis* survival, some of which might represent novel intervention targets against clonorchiasis if they are selective for the parasite. First, conserved genes linked to ubiquitin-mediated proteolysis, proposed to be involved in encystation, growth and reproduction of *C. sinensis* and the progression of cholangiocarcinoma (Huang, Liao, et al., 2013), and highly expressed in adults and other developmental stages of *C. sinensis* (Huang, Liao, et al., 2013; Yoo et al., 2011), might represent such target candidates. Second, invariable genes linked to the glucogon signalling pathway, including the gene coding for phosphoenolpyruvate carboxykinase, could be other candidates. The latter protein is one of the key enzymes regulating glyconeogenesis (cf. Huang, Chen, et al., 2013; Li et al., 2020; Yu et al., 2021) and is highly expressed in all tissues of the adult worm, including the oral sucker and muscles (Huang, Chen, et al., 2013). Third, a myophilin-like protein, which is associated with the regulation of smooth muscle contraction, and found to be highly expressed in the oral sucker and pharynx of *C. sinensis* (see Huang et al., 2012; Huang, Chen, et al., 2013), might represent a potential vaccine or drug target in this parasite, as proposed previously (Huang et al., 2012).

In conclusion, the discovery of marked genetic differentiation within *C. sinensis* in Far East Asia indicates the existence of biologically distinct population variants or, potentially, cryptic species. Such genetic variants might have distinct host preferences or affiliations, transmission patterns, pathogenicities and/or responsiveness to anthelmintic treatment (with praziquantel)—all worthy areas of future investigation. The present findings emphasize the importance of not making assumptions about genetic homogeneity of a morphospecies, and the relevance of using genomic-informatic tools to comprehensively investigate the genetic composition of parasite

populations. Although applied here to *C. sinensis*, the practical workflow established here for the genome-wide analysis of nucleotide variability in protein-coding genes should be applicable to a broad range of trematodes and other parasites for which high-quality nuclear genomes and curated data sets are available.

AUTHOR CONTRIBUTIONS

The research was designed and performed by L.K., P.K.K., N.D.Y. and R.B.G.; U.S., X.-Q.Z., I.H., B.Y., J.L.F., G.N.C. and R.B.G. provided resources for this research, and L.K., P.K.K., N.D.Y. and R.B.G. interpreted the data and wrote the article with input from all authors.

ACKNOWLEDGEMENTS

We thank Prof. Chun-Ren Wang, College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, and Prof. Ming-Yuan Liu, Key Laboratory of Zoonosis Research, Ministry of Education, Institute of Zoonosis, College of Veterinary Medicine, Jilin University, China, for providing some of the *C. sinensis* specimens used in this study. Dr V. V. Besprozvannykh is acknowledged for assistance with the collection of *C. sinensis* specimens from Russia. Research funding from the Australian Research Council (to Robin B. Gasser, Neil D. Young and Pasi K. Korhonen; LP180101085, LP180101334 and LP190101209) and from the Estonian Ministry of Education and Research (grant no. PRG1209 to Urmas Saarma) is gratefully acknowledged. Xing-Quan Zhu was supported by the Fund for Shanxi "1331 Project" (grant no. 20211331-13) and the Special Research Fund of Shanxi Agricultural University for High-level Talents (grant no. 2021XG001). The funders had no role in study design, data collection, analyses, decision to publish and/or writing of the manuscript. Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Aligned sequence read data and metadata have been deposited in the SRA (BioProject PRJNA860012).

BENEFIT-SHARING STATEMENT

Benefits from this research accrue via the sharing of our data and results on public databases.

ORCID

Pasi K. Korhonen https://orcid.org/0000-0002-9957-4674
Neil D. Young https://orcid.org/0000-0001-8756-229X
Robin B. Gasser https://orcid.org/0000-0002-4423-1690

REFERENCES

Anari, M., Stroehlein, A. J., Hall, R. S., Chang, B. C. H., Gasser, R. B., & Young, N. D. (2020). Expanded complement of Niemann-Pick type C2-like protein genes in *Clonorchis sinensis* suggests functions beyond sterol binding and transport. *Parasites & Vectors*, *13*(1), 38. https://doi.org/10.1186/s13071-020-3910-0

Bayne, C. J., & Grevelding, C. G. (2003). Cloning of *Schistosoma mansoni* sporocysts in vitro and detection of genetic heterogeneity among individuals within clones. *The Journal of Parasitology*, *89*(5), 1056–1060. https://doi.org/10.1645/GE-3186RN

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, *27*(8), 1767–1780. https://doi.org/10.1093/nar/27.8.1767

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, *18*(1), 237–268.

Chelomina, G. N. (2018). Clonorchis. In D. Liu (Ed.), *Handbook of food-borne diseases* (pp. 723–736). CRC Press.

Chelomina, G. N., Tatonova, Y. V., Hung, N. M., & Ngo, H. D. (2014). Genetic diversity of the Chinese liver fluke *Clonorchis sinensis* from Russia and Vietnam. *International Journal for Parasitology*, *44*(11), 795–810. https://doi.org/10.1016/j.ijpara.2014.06.009

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767–1771. https://doi.org/10.1093/nar/gkp1137

Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: Recommendations 1984. *Nucleic Acids Research*, *13*(9), 3021–3030.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*(1), 214. https://doi.org/10.1186/1471-2148-7-214

Ershov, N. I., Mordvinov, V. A., Prokhortchouk, E. B., Pakharukova, M. Y., Gunbin, K. V., Ustyantsev, K., Genaev, M. A., Blinov, A. G., Mazur, A., Boulygina, E., Tsygankova, S., Khrameeva, E., Chekanov, N., Fan, G., Xiao, A., Zhang, H., Xu, X., Yang, H., Solovyev, V., … Skryabin, K. G. (2019). New insights from *Opisthorchis felineus* genome: Update on genomics of the epidemiologically important liver flukes. *BMC Genomics*, *20*(1), 399. https://doi.org/10.1186/s12864-019-5752-8

Geyer, C. J. (1991). *Markov-Chain Monte-Carlo Maximum-Likelihood* (pp. 156–163). Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface, Interface Foundation, Fairfax, Virginia.

Guo, L., Zhang, S., Rubinstein, B., Ross, E., & Sánchez Alvarado, A. (2016). Widespread maintenance of genome heterozygosity in *Schmidtea mediterranea*. *Nature Ecology and Evolution*, *1*, 19. https://doi.org/10.1038/s41559-016-0019

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.2307/2334940

Hong, S. T., Yoon, K., Lee, M., Seo, M., Choi, M. H., Sim, J. S., Choi, B. I., Yun, C. K., & Lee, S. H. (1998). Control of clonorchiasis by repeated praziquantel treatment and low diagnostic efficacy of sonography. *The Korean Journal of Parasitology*, *36*(4), 249–254. https://doi.org/10.3347/kjp.1998.36.4.249

Huang, Y., Chen, W., Wang, X., Liu, H., Chen, Y., Guo, L., Luo, F., Sun, J., Mao, Q., Liang, P., Xie, Z., Zhou, C., Tian, Y., Lv, X., Huang, L., Zhou, J., Hu, Y., Li, R., Zhang, F., … Yu, X. (2013). The carcinogenic liver fluke, *Clonorchis sinensis*: New assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes. *PLoS One*, *8*(1), e54732. https://doi.org/10.1371/journal.pone.0054732

Huang, Y., Li, W., Huang, L., Hu, Y., Chen, W., Wang, X., Sun, J., Liang, C., Wu, Z., Li, X., Xu, J., & Yu, X. (2012). Identification and characterization of myophilin-like protein: A life stage and tissue-specific antigen of *Clonorchis sinensis*. *Parasitology Research*, *111*(3), 1143–1150. https://doi.org/10.1007/s00436-012-2946-2

Huang, Y., Liao, H., Li, W., Hu, Y., Huang, L., Wang, X., Sun, J., Chen, W., Deng, C., Liang, C., Wu, Z., Li, X., Xu, J., & Yu, X. (2013). Identification, sequence analysis and characterization of *Clonorchis sinensis* ubiquitin. *Experimental Parasitology*, *133*(1), 62–69. https://doi.org/10.1016/j.exppara.2012.10.015

International Agency for Research in Cancer (IARC) Working Group on the Evaluation of Carcinogenic Risks to Humans. (2012). Biological agents. Volume 100 B. A review of human carcinogens. *IARC Monographs on the Identification of Carcinogenic Hazards to Humans*, *100*(Pt B), 1–441.

Jolliffe, I. T. (1986). Principal component analysis and factor analysis. In *Principal component analysis* (pp. 115–128). Springer.

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Kim, T.-S., Pak, J. H., Kim, J.-B., & Bahk, Y. Y. (2016). *Clonorchis sinensis*, an oriental liver fluke, as a human biological agent of cholangiocarcinoma: A brief review. *BMB Reports*, *49*(11), 590–597. https://doi.org/10.5483/bmbrep.2016.49.11.109

Kinkar, L., Korhonen, P. K., Wang, D., Zhu, X.-Q., Chelomina, G. N., Wang, T., Hall, R. S., Koehler, A. V., Harliwong, I., Yang, B., Fink, J. L., Young, N. D., & Gasser, R. B. (2020). Marked mitochondrial genetic variation in individuals and populations of the carcinogenic liver fluke *Clonorchis sinensis*. *PLoS Neglected Tropical Diseases*, *14*(8), e0008480. https://doi.org/10.1371/journal.pntd.0008480

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Korhonen, P. K., Shaban, B., Faux, N. G., Kinkar, L., Chang, B. C. H., Wang, D., Yang, B., Young, N. D., & Gasser, R. B. (2022). "Escalibur"–A practical pipeline for the de novo analysis of nucleotide variation in nonmodel eukaryotes. *Molecular Ecology Resources*, *22*(5), 2120–2126. https://doi.org/10.1111/1755-0998.13600

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, S., Chen, X., Zhou, J., Xie, Z., Shang, M., He, L., Liang, P., Chen, T., Mao, Q., Liang, C., Li, X., Huang, Y., & Yu, X. (2020). Amino acids serve as an important energy source for adult flukes of *Clonorchis sinensis*. *PLoS Neglected Tropical Diseases*, *14*(4), e0008287. https://doi.org/10.1371/journal.pntd.0008287

Lin, R.-Q., Tang, J.-D., Zhou, D.-H., Song, H.-Q., Huang, S.-Y., Chen, J.-X., Chen, M.-X., Zhang, H., Zhu, X.-Q., & Zhou, X.-N. (2011). Prevalence of *Clonorchis sinensis* infection in dogs and cats in subtropical southern China. *Parasites & Vectors*, *4*(1), 180. https://doi.org/10.1186/1756-3305-4-180

Lun, Z.-R., Gasser, R. B., Lai, D.-H., Li, A.-X., Zhu, X.-Q., Yu, X.-B., & Fang, Y.-Y. (2005). Clonorchiasis: A key foodborne zoonosis in China. *The Lancet Infectious Diseases*, *5*(1), 31–41. https://doi.org/10.1016/S1473-3099(04)01252-6

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics*, *7*, 29–59. https://doi.org/10.1146/annurev.genom.7.080505.115623

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Moore, W. S. (1995). Inferring phylogenies from mtdna variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, *49*(4), 718–726. https://doi.org/10.1111/j.1558-5646.1995.tb02308.x

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics*, *34*(5), 867–868. https://doi.org/10.1093/bioinformatics/btx699

Qian, M.-B. (2014). Clonorchiasis control: Starting from awareness. *Infectious Diseases of Poverty*, *3*, 33. https://doi.org/10.1186/2049-9957-3-33

Qian, M.-B., Chen, Y.-D., Liang, S., Yang, G.-J., & Zhou, X.-N. (2012). The global epidemiology of clonorchiasis and its relation with cholangiocarcinoma. *Infectious Diseases of Poverty*, *1*, 4. https://doi.org/10.1186/2049-9957-1-4

Qian, M.-B., Utzinger, J., Keiser, J., & Zhou, X.-N. (2016). Clonorchiasis. *The Lancet*, *387*(10020), 800–810. https://doi.org/10.1016/S0140-6736(15)60313-0

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramm, S. A. (2017). Exploring the sexual diversity of flatworms: Ecology, evolution, and the molecular biology of reproduction. *Molecular Reproduction and Development*, *84*(2), 120–131. https://doi.org/10.1002/mrd.22669

Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11*(1), 7. https://doi.org/10.1186/1471-2105-11-7

Tang, Z.-L., Huang, Y., & Yu, X.-B. (2016). Current status and perspectives of *Clonorchis sinensis* and clonorchiasis: Epidemiology, pathogenesis, omics, prevention and control. *Infectious Diseases of Poverty*, *5*, 71. https://doi.org/10.1186/s40249-016-0166-1

Tatonova, Y. V., Chelomina, G. N., & Besprosvannykh, V. V. (2012). Genetic diversity of nuclear ITS1-5.8S-ITS2 rDNA sequence in *Clonorchis sinensis* Cobbold, 1875 (Trematoda: Opisthorchidae) from the Russian Far East. *Parasitology International*, *61*(4), 664–674. https://doi.org/10.1016/j.parint.2012.07.005

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*, 57–86.

Wang, D., Korhonen, P. K., Gasser, R. B., & Young, N. D. (2018). Improved genomic resources and new bioinformatic workflow for the carcinogenic parasite *Clonorchis sinensis*: Biotechnological implications. *Biotechnology Advances*, *36*(4), 894–904. https://doi.org/10.1016/j.biotechadv.2018.02.008

Wang, D., Young, N. D., Korhonen, P. K., & Gasser, R. B. (2018). *Clonorchis sinensis* and clonorchiasis: The relevance of exploring genetic variation. *Advances in Parasitology*, *100*, 155–208. https://doi.org/10.1016/bs.apar.2018.03.006

Wang, X., Chen, W., Huang, Y., Sun, J., Men, J., Liu, H., Luo, F., Guo, L., Lv, X., Deng, C., Zhou, C., Fan, Y., Li, X., Huang, L., Hu, Y., Liang, C., Hu, X., Xu, J., & Yu, X. (2011). The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. *Genome Biology*, *12*(10), R107. https://doi.org/10.1186/gb-2011-12-10-r107

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Williams, R. B. H., Chan, E. K. F., Cowley, M. J., & Little, P. F. R. (2007). The influence of genetic variation on gene expression. *Genome Research*, *17*(12), 1707–1716. https://doi.org/10.1101/gr.6981507

World Health Organization (WHO). (2020). *Ending the neglect to attain the sustainable development goals: A road map for neglected tropical diseases 2021–2030: Overview*. World Health Organization.

Yangco, B. G., De Lerma, C., Lyman, G. H., & Price, D. L. (1987). Clinical study evaluating efficacy of praziquantel in clonorchiasis. *Antimicrobial Agents and Chemotherapy*, *31*(2), 135–138.

Yoo, W. G., Kim, D.-W., Ju, J.-W., Cho, P. Y., Kim, T. I., Cho, S.-H., Choi, S.-H., Park, H.-S., Kim, T.-S., & Hong, S.-J. (2011). Developmental transcriptomic features of the carcinogenic liver fluke, *Clonorchis sinensis*. *PLoS Neglected Tropical Diseases*, *5*(6), e1208. https://doi.org/10.1371/journal.pntd.0001208

Young, N. D., Campbell, B. E., Hall, R. S., Jex, A. R., Cantacessi, C., Laha, T., Sohn, W.-M., Sripa, B., Loukas, A., Brindley, P. J., & Gasser, R. B. (2010). Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. *PLoS Neglected Tropical Diseases*, *4*(6), e719. https://doi.org/10.1371/journal.pntd.0000719

Young, N. D., Stroehlein, A. J., Kinkar, L., Wang, T., Sohn, W.-M., Chang, B. C. H., Kaur, P., Weisz, D., Dudchenko, O., Aiden, E. L., Korhonen, P. K., & Gasser, R. B. (2021). High-quality reference genome for *Clonorchis sinensis*. *Genomics*, *113*(3), 1605–1615. https://doi.org/10.1016/j.ygeno.2021.03.001

Yu, S., Meng, S., Xiang, M., & Ma, H. (2021). Phosphoenolpyruvate carboxykinase in cell metabolism: Roles and mechanisms beyond gluconeogenesis. *Molecular Metabolism*, *53*, 101257. https://doi.org/10.1016/j.molmet.2021.101257

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kinkar, L., Korhonen, P K., Saarma, U., Wang, T., Zhu, X.-Q., Harliwong, I., Yang, B., Fink, J L., Wang, D., Chang, B C H., Chelomina, G N., Koehler, A V., Young, N D., & Gasser, R B. (2023). Genome-wide exploration reveals distinctive northern and southern variants of *Clonorchis sinensis* in the Far East. *Molecular Ecology Resources*, *23*, 833–843. https://doi.org/10.1111/1755-0998.13760