Evolving in the darkness: phylogenomics of *Sinocyclocheilus* cavefishes highlights recent diversification and cryptic diversity

Tingru Mao[1,†], Yewei Liu[1,†], Mariana M. Vasconcellos[2], Marcio R. Pie[3], Gajaba Ellepola[1], Chenghai Fu[1], Jian Yang[4], Madhava Meegaskumbura[1,*]

[1]Guangxi Key Laboratory for Forest Ecology and Conservation, College of Forestry, Guangxi University, Nanning, Guangxi, P.R.C.

[2]Programa de Pós-Graduação em Ecologia. Universidade Federal do Rio de Janeiro. Rio de Janeiro - RJ. Brazil

[3]Departamento de Zoologia, Universidade Federal do Paraná, Brazil, 81531-980

[4]Key Laboratory of Environment Change and Resource Use, Beibu Gulf, Nanning Normal University, Nanning, Guangxi, P.R.C.

[†] These authors contributed equally

Corresponding author *E-mail*: madhava_m@mac.com

ABSTRACT

Troglomorphism— any morphological adaptation enabling life to the constant darkness of caves, such as loss of pigment, reduced eyesight or blindness, over-developed tactile and olfactory organs—has long intrigued biologists. However, inferring the proximate and ultimate mechanisms driving the evolution of troglomorphism in freshwater fish requires a sound understanding of the evolutionary relationships between surface and troglomorphic lineages. We use Restriction Site Associated DNA Sequencing (RADseq) to better understand the evolution of the *Sinocyclocheilus* fishes of China. With a remarkable array of derived troglomorphic traits, they comprise the largest cavefish diversification in the world, emerging as a multi-species model system to study evolutionary novelty. We sequenced a total of 120 individuals throughout the *Sinocyclocheilus* distribution. The data comprised a total of 646,497bp per individual, including 4378 loci and 67,983 SNPs shared across a minimum of 114 individuals at a given locus. Phylogenetic analyses using either the concatenated RAD loci (RAxML) or the SNPs under a coalescent model (SVDquartets, SNAPP) showed a high degree of congruence with similar topologies and high node support (> 95 for most nodes in the phylogeny). The major clades recovered conform to a pattern previously established using Sanger-based mt-DNA sequences, with a few notable exceptions. We now recognize six major clades in this group, elevating the blind cavefish *S. tianlinensis* and the micro-eyed *S. microphthalmus* as two new distinct clades due to their deep divergence from other clades. PCA plots of the SNP data also supports the recognition of six major clusters of species congruent with the identified clades based on the spatial arrangement and overlap of the species in the PC space. A Bayes factor delimitation (BFD) analysis showed support for 21 species, recognizing 19 previously described species and two putative new cryptic ones. Two species whose identities were previously disputed, *S. furcodorsalis* and *S. tianeensis,* are supported here as distinct species. In addition, our multi-species calibrated tree in SNAPP suggests that the genus *Sinocyclocheilus* originated around 10.5 Mya, with most speciation events occurring in the last 2 Mya, likely favored by the uplift of the Qinghai-Tibetan Plateau and cave occupation induced by climate-driven aridification during this period. These results provide a firm basis for future comparative studies on the evolution of *Sinocyclocheilus* and its adaptations to cave life.

KEYWORDS: Phylogenomics, RADseq, diversification, cavefish, species delimitation, introgression

1.INTRODUCTION

Many animal lineages across the world, ranging from flatworms (Leal-Zanchet et al., 2014) to fish and amphibians (Hutchison, 1958), have evolved to live in caves. In order to occupy cave subterranean habitats, each of these lineages had to adapt to extreme conditions of low availability of food, oxygen, and light, leading to the repeated acquisition of one or more adaptations, including elongated appendages, lowered metabolism, specialized sensory systems, and loss of eyes and pigmentation (Jeffery, 2019). The concerted evolution of these traits across different taxa, giving rise to convergent troglomorphic forms, comprise a great example of the process of natural selection in response to cave-associated selective regimes (Borowsky, 2010; Jeffery et al., 2010; Klaus et al., 2013; Porter et al., 2003). Though large troglomorphic radiations are rare – mainly because of the limited extent of cave habitats – the cyprinid fish genus *Sinocyclocheilus* shows an exceptional diversity of cave adaptations including at least three independent origins of cave-adapted phenotypes (Mao et al., 2021).

With 75 known species, *Sinocyclocheilus* includes the largest radiation of cavefish in the world, which diversified across Karstic habitats associated with the Li River in the Guangxi, Guizhou, and Yunnan provinces of China (Jiang et al., 2019; Zhao and Zhang, 2009). The troglomorphic adaptations of *Sinocyclocheilus* include the development of a specialized sensory system, such as degeneration or complete loss of eyes, degeneration of scales, loss of pigmentation, expansion of pectoral fins, evolution of 'horns' and enhancement of the neuromast system (He et al., 2013; Li et al., 2020; Ma et al., 2020; Meng et al., 2013). This wealth of troglomorphic adaptations makes *Sinocyclocheilus* an important multi-species model system to study evolutionary novelty in response to selection (Chen et al., 2009; Meng et al., 2013; Xiao et al., 2005; Yang et al., 2016). Yet, this requires a robust time-calibrated phylogeny for this group including several independent markers across their genome, which is currently missing.

Our current understanding of the phylogenetic relationships within *Sinocyclocheilus,* based solely on mitochondrial DNA (mtDNA), revolves around the recognition of four main clades, A, B, C, and D (Mao et al., 2021), sometimes also referred to as the "jii", "angularis", "cyphotergous", and" tingi" groups (Li and He, 2009; Xiao et al., 2005; Zhao and Zhang, 2009). These are sequential clades rather than reciprocally monophyletic units. The earliest emerging clade (Clade A) is restricted to Guangxi, at the eastern fringes of the genus distribution. Clades B and C have overlapping distributions restricted to the middle of the genus distribution, and species of Clade D are found mostly in the lotic habitats associated with hills to the west. In addition, Clade B encompasses most species with extensive troglomorphic traits such as the complete loss of eyes (blind) and well-formed forehead protrusions (horns). The genus is thought to have originated during the Miocene-Pliocene with cave occupation taking place predominantly during the Pliocene-Pleistocene transition in response to an uplift event and aridification of the Qinghai-Tibetan Plateau (Mao et al., 2021).

Nearly all of the molecular studies to date on the phylogenetic relationships within *Sinocyclocheilus* have been based on mtDNA (Chen et al., 2018; Jiang et al., 2019; Liang et al., 2011), limiting some of the diversification-scale analyses as well as insights into the population-level processes that shape this remarkable radiation. This is an important limitation, given that the uniparental nature of mtDNA inheritance might obscure important events of admixture in the history of *Sinocyclocheilus* while hampering the inference of species limits given that it represents only a single genetic locus. With the advent of next-generation sequencing, particularly highly efficient methods for recovering thousands of orthologous loci such as Restriction-site associated DNA sequencing (Cariou et al., 2013), we can infer the relationships within *Sinocyclocheilus* with higher accuracy*,* as well as its species limits, ancestral admixture, and diversification events.

The main goals of our study were: (1) to infer the evolutionary relationships and divergence times within *Sinocyclocheilus* through phylogenomic methods; (2) to assess the level of genetic support for the currently recognized species in the genus using Bayesian species delimitation methods; and (3) to investigate possible ancestral admixture events among *Sinocyclocheilus* species. In short, we were able to build a well-resolved tree for the genus despite recognizing some introgression events, to confirm the distinctiveness of 19 previously described species, and to recognize two additional cryptic species. Finally, we estimate that the most recent common ancestor (MRCA) of *Sinocyclocheilus* originated around 10.5 Mya, which is considerably older than previous estimates based on mitochondrial data alone.

2. METHODS

2.1 Taxon sampling and laboratory work

To elucidate the phylogenetic relationships of *Sinocyclocheilus* using multiple nuclear genomic markers, we carried out a detailed phylogenomic analysis of the group. We collected samples of 120 individuals in Guangxi, Yunnan and Guizhou provinces of China, including 19 previously recognized species and 2 unidentified species of *Sinocyclocheilus* (Fig. S1). Given the rarity of these species, only fin-clips were collected from all individuals and frozen immediately at -85 °C until DNA extraction. Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA) following the manufacturer's protocols. Electrophoresis was performed to ensure DNA integrity of the samples prior to genomic library preparation.

We employed a RAD sequencing protocol (Baird et al., 2008), which involves the use of a single restriction enzyme to cut genomic DNA at specific sites, using molecular barcodes to identify each individual and build a genomic library of short fragments distributed across the entire genome of all individuals. Each DNA sample was digested using the EcoRI enzyme and ligated to a P1 Illumina adapter at the compatible ends of the fragments. This adapter allows forward amplification with Illumina sequencing primers containing a 6-bp long nucleotide barcode for sample identification. The barcoded adapter-ligated fragments from all individuals were subsequently pooled, randomly sheared, and size-selected. DNA was then ligated to a second P2 adapter, a Y-shaped adapter with divergent ends. Finally, fragment sizes of 200-400 bp and 400-600 bp were isolated for library construction using a MinElute Gel Extraction kit (Qiagen). We quantified DNA concentration in a Qubit2.0, and checked the quality (insert size) of genomic libraries in an Agilent 2100. qPCR was also performed to detect the effective concentration of libraries (if > 2nM) in the appropriate insert size. Final genomic libraries were then sequenced on an Illumina HiSeq2500 platform generating 125-bp paired-end reads.

2.2 RADseq data assembly

Our pre-processing bioinformatics of the sequenced genomic library consisted of the following: first, the raw reads of fastq format were filtered through in-house scripts to optimize read number and to reduce artifacts within the data assembly. In this step, clean reads were obtained by removing low-quality reads and reads containing adapter sequence or poly-N from the raw data. In addition, the number of reads from each RAD site were tracked, and RAD sequences above a certain threshold were removed since highly repetitive sequences most likely represent paralogs. Single mismatch derivatives of these highly repetitive RAD sequences were also removed. Finally, low frequency RAD sequences below a threshold (of ≤5) were also removed from further analysis, as the associated paired-end reads would lack sufficient coverage for accurate genotyping (SNP calling).

We processed the clean RADseq reads using the ipyrad 0.9.56 pipeline assembly (Eaton and Overcast, 2020), which is well suited for downstream phylogenetic analyses (Leaché et al., 2015). Since there was no suitable reference genome, we used a *de novo* method to assemble the R1 clean reads. During demultiplexing, restriction sites were trimmed from all reads. All individuals had deep sequencing and high-quality assembly, hence all individuals were retained for downstream analyses. Given the genus *Sinocyclocheilus* is tetraploid (Li and Guo, 2020), we increased the depth of coverage of retained loci to 10x, for accurate genotyping and unbiased heterozygosity. Thus, the maximum number of alleles in the individual consensus sequence was set to 4, to allow more allele combinations within each individual. We set the clustering threshold to 90% similarity and allowed trimmed reads of at least 75bp to proceed in the assembly. In addition, datasets with various minimum numbers of individuals per locus were filtered (96, 102, 108, 114 and 120, corresponding to 80–100% of individual coverage). Other parameters in ipyrad were set to default values.

Demultiplexed, unfiltered reads for the RAD data pertaining to this study can be accessed through the NCBI GenBank Short Read Archive (PRJNA764266).

2.3 Phylogenomic analyses

We used the concatenated sequences from all loci of the m114 matrix (8.44% missing data), corresponding to at least 95% individual coverage across loci to infer phylogenetic relationships among all samples. Phylogenomic relationships of all *Sinocyclocheilus* samples were inferred based on the maximum likelihood analysis of a matrix with 1,560,414 characters under a GTRCAT model (Stamatakis, 2014). We also inferred the species tree using the unlinked SNP dataset of the m114 matrix (with a single SNP sampled from each locus) in SVDquartets (Chifman and Kubatko, 2014) implemented in PAUP* 4.0a (Wilgenbusch and Swofford, 2003). SVDquartets estimates the relationships among taxa under the coalescent model by inferring splits among quartets of randomly sampled taxa. All possible taxa quartets (8,214,570 quartets) were evaluated and node support was estimated with 100 bootstrap replicates. The species tree inferred in SVDquartets can assist in identifying incomplete lineage sorting (ILS) or introgression among species in nodes with low support. All phylogenies were visualized using FigTree v1.4.3.

2.4 Population cluster analyses

We used a principal component analysis (PCA) to downscale genomic differences among individuals into the first two principal components and visualize the differences between populations. The PCA was based on the .snps.hdf file generated by ipyrad after assembling the data using the ipyrad.pca tool (Eaton and Overcast, 2020). Prior to the analysis, we performed the following treatments: (1) Assigning samples to populations based on the phylogenetic relationships using the imap dictionary. (2) Filtering SNP data using the minmap dictionary to ensure that 50%, 60%, 70%, 80%, 90% of samples have data in each group. (3) Inputting missing data using the "sample" algorithm. Each RAD locus was randomly subsampled to a single SNP to reduce the effect of linkage on the results. After filtering, 3,459, 2,942, 1,993, 987, and 165 SNPs were included in the analysis.

2.5 Species delimitation method

We performed Bayes factor species delimitation analyses using BFD* (Leaché et al., 2014). This method computes the marginal likelihood of species trees inferred with the software SNAPP using the Path Sampling approach in BEAST2 (Bouckaert et al., 2014). The method can be used to compare different species delimitation models using genome-wide SNP data in a multispecies coalescent framework. Due to computational limitations, we did not analyze all 120 samples simultaneously. We conducted three independent analyses of species delimitation (Table 1): the first including *S.* cf. *guanyangensis* with the other three closely related taxa in clade A (26 samples), the second including *S.* cf. *longibarbatus* with the other four closely-related taxa in clade C (28 samples), and the third including *S. furcodorsalis* with the other 4 closely-related taxa in clade B (26 samples). The data were filtered to include no missing data among selected samples (Table 1). The parameters for SNAPP were set following Leaché and Bouckaert (2018) for the BFD* analyses (Leaché and Bouckaert, 2018). We conducted a path sampling for a total of 48 steps (MCMC length = 100,000, pre-burnin = 10,000) to calculate the marginal likelihood estimation (MLE) for all models, ranking them by comparing the size of the Bayes Factors across different models.

2.6 Divergence time estimation

We inferred divergence times for the *Sinocyclocheilus* using the MSC model in SNAPP, building a species tree of unlinked SNPs following Stange et al (2018), as this algorithm can compensate for ascertainment bias introduced by the exclusion of invariable sites (Bryant et al., 2012; Stange et al., 2018). As running the complete data set was computationally challenging, we selected a single representative per species (21 terminal clades) and reassembled them in ipyrad (2,602 unlinked SNPs for 21 individuals). We then used the Ruby script snapp_prep.rb to prepare the XML input for SNAPP (available at https://github.com/mmatschiner/snapp_prep). For time calibration, we used the time inferred by Liang et al (2011) to the split between *S. donglanensis* and its sister group *S. lingyunensis* at approximately 1.32 Mya (SD:1.02, 95%CI: 0.08-3.91) (Liang et al., 2011). We then ran an independent SNAPP analysis with 20 million MCMC generations, sampling at every 2,000 steps. We checked for stationarity and convergence of chains in TRACER v1.7.1 (ESS>200). A maximum clade credibility tree was generated with TreeAnnotator discarding the first 10% of each MCMC chain as a burn-in. The program FigTree v.1.4.3 was used to visualize the summary tree.

2.7 Tests of introgression

To understand the causes of conflicting phylogenetic signals in the species and gene trees of *S. tianlinensis*, *S. yishanensis*, and *S. macrophthalmus*, we used Treemix from the ipyrad analysis toolkit inferring instances of ancestral gene flow (Pickrell and Pritchard, 2012). Prior to the Treemix analysis, the dataset was processed and filtered in the same way as in the principal component analysis, obtaining 3,459 SNPs. We tested the number of migration edges (m) in the range of 1 to 10, estimating their likelihood score to determine the appropriate value for migration edge in this data. Adding additional admixture edges will always improve the likelihood score, but with diminishing returns as you add additional edges that explain little variation in the data (Kim et al., 2021; Popovic et al., 2020).

3. RESULTS

3.1. RADseq dataset

The RADseq genomic libraries of the 120 individuals of *Sinocyclocheilus* yielded approximately 9.8 million reads per individual on average, of which 99.91% were retained after the filtering step of the assembly (supplementary table S1). We selected the parameter combination Sino_m114 (supplementary table S2) for phylogenetic analyses as it had the optimal combination of loci-clustering parameters for *Sinocyclocheilus* that maximized the fraction of variable sites that were phylogenetically informative. The data comprised a total of 646,497bp, including 4,378 loci and 67,983 SNPs shared across at least 114 individuals (95% individual coverage at a given locus), of which 61,023 SNPs were parsimony informative for phylogenetic analyses (Sino_m114; supplementary table S2).

3.2 Phylogenetic reconstruction of RADseq data

The maximum likelihood analysis of the concatenated RADseq loci in RAxML and the coalescent-based SVDquartets analysis of unlinked SNP data recovered similar phylogenetic relationships among the 19 known *Sinocyclocheilus* species with a high degree of support (> 95) for most nodes (Fig. 1, Fig. 2). As expected, the bootstrap support values at a few nodes of the tree obtained by the coalescent SVDquartets analysis were slightly lower than those obtained by the concatenated analysis (Fig. 1). The two nodes with low bootstrap support in the SVDquartets phylogeny were: the one separating *S. macrophthalmus* and *S. yishanensis* (= 74), and that corresponding to the MRCA of *S. macrophthalmus, S. yishanensis, S. lingyunensis, S. donglanensis* (= 55) (Fig. 2), which suggests some uncertainty in the placement of *S. macrophthalmus* and *S. yishanensis*. Our phylogenetic analyses (Figs. 1-2, A–F) recovered *Sinocyclocheilus* as consisting of six major clades, four of which has been recognized before (Clades A, B, C, and D – Mao et al. 2021). *Sinocyclocheilus tianlinensis,* recognized as a new Clade E, is recovered as the sister species to Clades B, C, D, and F, which is not consistent with the topology of the mtDNA phylogeny (Mao et al., 2021). Likewise, *S. microphthalmus*, recognized as a new Clade F, is now recovered as the sister species to Clades C and D, which is also not congruent with previous phylogenies. Additional incongruences to previous mt-DNA based studies in the Clade C include: *S. longibarbatus* recovered as the sister species to *S. xunlensis*, and *S. yishanensis* as the sister species to *S. macrophthalmus* (Mao et al., 2021). Interestingly, one incongruence also occurred across our trees inferred by different methods: *S. tianlinensis* had a different relationship with a somewhat low bootstrap value in the SNAPP tree (Fig. 3). This suggests conflicting signal, possibly due to introgression. Other than that, relationships across all analyses were congruent for all the remaining species.

3.3 Population clustering

We used PCA plots to summarize and visualize in two dimensions the genomic differences across all samples of *Sinocyclocheilus*. The first two principal components explained between 32.2% and 24.9% of the variation using different thresholds for data completeness (50% – 90%) (Fig. 4). The first two axes showed that samples clustered into six subgroups congruent with the major clades identified in our phylogenies.

3.4 Species delimitation

The results for all models tested with BFD* method are summarized in Table 1. Models A1 and C1 support the recognition of *S .cf. guanyangensis* (MLE= -2043.41) and *S.* cf. *longibarbatus* (MLE= -1581.81) as separate species. In addition, whether *S. furcodorsalis* and *S. tianeensis* are the same species is also controversial (Liang et al., 2011; Zhao and Zhang, 2009). However, using the BFD* method, the model representing the current taxonomy showed a high Bayes Factor value (MLE= -1564.99), suggesting that *S. furcodorsalis* and *S. tianeensis* are indeed distinct species, supporting the current taxonomy.

3.5 Divergence times estimates

Our time-calibrated tree in SNAPP suggests that the crown age for the genus *Sinocyclocheilus* is around 10.5 My old. In an MSC model, *S. tianlinensis* and Clade B show a sister-group relationship (with probability of 0.77). In Clade C, *S. yishanensis* and *S. macrophthalmus* (probability 1.00), *S. lingyunensis* and *S. donglanensis* (probability 0.77) formed sub-clades. (Fig. 3). Compared with the concatenated ML and SVDquartets trees, the SNAPP tree showed a different topology with 5 major clades, including *S. tianlinensis* in Clade B.

3.6 Treemix analysis

The likelihoods for different migration boundaries increased steadily as more edges are included in the model (from 0 to 10). However, the likelihood did not improve significantly when the number of migration edges increased above 5. Treemix analyses of m=0–5 suggests the possibility of ancestral gene flow in *S. tianlinensis* and the ancestral species of *S. furcodorsalis* and *S. tianeensis.* Moreover, the possibility of gene flow also exists in the ancestors of *S. yishanensis*, *S. macrophthalmus* and *S. lingyunensis*, *S. donglanensis* (Fig. 5).

4. DISCUSSION

4.1 Phylogeny and cryptic species of *Sinocyclocheilus*

A total of 75 species of *Sinocyclocheilus* fish from China has been described up to now (Jiang et al., 2019), mainly based on morphology and mtDNA data. Nevertheless, some authors have questioned the validity of a few of those species (Liang et al., 2011; Zhao and Zhang, 2009) mainly due to their morphological similarities and incongruences across mt-DNA based phylogenies. For example, the validity of *S. furcodorsalis* and *S. tianeensis* has been questioned by several authors due to their close morphological resemblance (Liang et al., 2011). Such controversies have hindered drawing general conclusions concerning the evolution, ecology, conservation, and biogeography of *Sinocyclocheilus.* Therefore, this genus is in urgent need of a resolved a phylogeny and a clear taxonomy to support robust comparative analyses of the evolution of cave-adaptations in this group.

To this end, we tested the species boundaries in *Sinocyclocheilus* using RAD sequencing-based molecular species delimitation, which had found wide utility in previous evolutionary research (Herrera et al., 2016; Leaché et al., 2014; Rancilhac et al., 2019; Razkin et al., 2016). In addition, *Sinocyclocheilus* had never been subjected to a rigorous molecular species delimitation analysis before. Furthermore, while morphological changes in *Sinocyclocheilus* occur rapidly, molecular divergence, especially in regions under neutral evolution, is assumed to be limited. Therefore, we presumed that mt-DNA based molecular delimitation would not reflect accurate phylogenetic relationships in this group as clearly as the current study, based on a genome-scale sampling that also includes regions under selection. This increased genomic sampling may reflect more accurately the species tree in which *Sinocyclocheilus* evolved their specialized morphology for cave dwelling. Nevertheless, the major conclusions based on the mtDNA phylogeny of Mao et al. (2021) - that blind species independently evolved at least three times - is supported by this RAD-marker based phylogeny as well (Mao et al., 2021).

Trees based on RAXML, SVDquartets and SNAPP were for the most part congruent comprising highly supported nodes with only two nodes on the species trees with low bootstrap support. These species tree methods (i.e. SVDquartets or SNAPP) analyze each SNP separately based on the coalescent theory, so they are expected to show lower bootstrap support whenever there is a conflicting signal among different SNPs in the dataset (Leaché et al., 2015). The tree generated with SNAPP also had a few inconsistencies with the other trees. In the SNAPP tree, *S. tianlinensis* is a sister species to Clade B while *S. donglanensis* and *S. lingyunensis* are sister species to Clade C. However, in the concatenated RAxML tree, *S. tianlinensis* is a sister species to Clades B, C, D, F, while *S. donglanensis*, *S. lingyunensis* are a sister clade to *S. macrophthalmus* and *S. yishanensis* in Clade C. We assume that this conflicting phylogenetic signal is either due to ancestral gene flow, introgression or incomplete lineage sorting. Treemix analysis suggests introgressive gene flow.

Previous evolutionary studies based on mitochondrial genes have shown that *Sinocyclocheilus* is divided into four major clades (Liang et al., 2011; Ma et al., 2019; Romero et al., 2009; Zhao and Zhang, 2009). However, the phylogenetic relationship obtained with RAD-seq analysis in the present study suggests that this group can be further divided into two additional branches, comprising six major clades (Fig. 1). The principal component analysis further supports this division by clustering the samples into six clusters which correspond to the six clades (Fig. 4). Previous mt-DNA based phylogenies (Mao et al., 2021) recognized *S. tianlinensis* and *S. microphthalmus* as species within clade B, but our phylogenomic results indicate that those two species actually belong to two independent clades due to their deep divergences in Clades B and C respectively.

Our results also support that *S. guanyangensis* and *S.* cf. *guanyangensis,* which resemble each other morphologically, are in fact two distinct species, the latter being a putative new species. The same is true for *S. longibarbatus* and *S.* cf. *longibarbatus* (a putative new species) as well. Field sampling indicated that these species pairs occur in different caves, but in close proximity to each other (see map in Fig. S1). Based on morphological and molecular data, it appears that different species of cave fish have repeatedly invaded different cave waters and acquired their own independent troglomorphic characteristics. This may be the case for these species as well. Geographical isolation followed by strong genetic drift among populations seems to be the main mechanism for the rapid formation of species in the genus *Sinocyclocheilus* (Ma et al., 2019; Yang et al., 2021; Zhao and Zhang, 2009). However, especially in small, isolated populations such as *S.* cf. *guanyangensis* and *S.* cf. *longibarbatus*, genetic differentiation and species formation may occur even more rapidly due to the lack of gene flow among isolated populations. This mechanism, occurring under strong selective pressure associated with cave dwelling, may accelerate and amplify the rapid evolution of adaptive traits in other species of *Sinocyclocheilus* as well, resulting in a particularly rich species diversity with frequent evolution to cave-adapted morphologies.

Our genome-wide SNP analyses identified two putative new species of *Sinocyclocheilus* out of the 120 individuals sampled across 21 localities. Considering that our study did not include samples for all 75 species described so far, we suspect that cryptic diversity in the group could be potentially higher, i.e., the species diversity of the *Sinocyclocheilus* complex may be vastly underestimated.

4.2 Divergence time estimation and gene flow

The age of the most recent common ancestor (MRCA) of *Sinocyclocheilus* fish, inferred by our RADseq data, was estimated to be around 10.5 Mya, similar to that inferred by previous mtDNA- based studies (Li et al., 2008; Liang et al., 2011; Mao et al., 2021). We also found that most divergence events occurred relatively recent in the history of the group (in the last 2 Mya) (Fig. 3). The *Sinocyclocheilus* MRCA possibly lived on the Yunnan-Guizhou Plateau during the late Tertiary. During the Quaternary, the Qinghai-Tibetan Plateau underwent an abrupt upturn, with major changes in the geological environment and a dramatic transformation of the Yunnan-Guizhou Plateau (Li and Fang, 1999; Shi et al., 1999). At the same time, global temperature began to fall and Northern Hemisphere ice caps grew (Hewitt, 2000; Ma et al., 2019; Svendsen et al., 2004). The MRCA of the *Sinocyclocheilus* radiation may have adapted to live in underground caves due to dramatic changes in the environment. The relatively late tectonic uplift of the Tibetan Plateau 3.6 Mya may have affected population dynamics of *Sinocyclocheilus* as well. For example, three species (*Sinocyclocheilus grahami*, *S. rhinocerous*, and *S. anshuiensis*) experienced two episodes of population declines during the two intense uplift phases (Qingzang movement: 3.6 Mya~1.7Mya, Kunhuang movement: 1.1Mya~0.6Mya) of the third tectonic uplift of the Qinghai-Tibet Plateau (Yang et al., 2016). Changes in population size and gene flow may be related to enhanced Asian monsoon and precipitation, with large-scale glacial activity resulting from these two phases considerably affecting diversification in this group. Therefore, the intense late tectonic activity in the Qinghai-Tibet Plateau, combined with environmental changes, could have promoted the rate of diversification in these cavefish during the last 2 My.

SNAPP is a multispecies coalescent method that assumes no gene flow among species. This may lead to inaccurate inference of topology in the presence of gene flow and may explain the incongruences among the SNAPP tree and the other trees. In the Treemix analysis, we found possible ancestral gene flow between some species of *Sinocyclocheilus*, which is likely responsible for the inconsistent phylogenetic signals described previously (Fig. 5). Subterranean river-capture events may have influenced this phenomenon. The geographical locations of the species showing possible ancestral genetic admixture are in close proximity and were possibly connected through underground networks, especially during the Pleistocene. With the upliftment of the Tibetan plateau during the early Pleistocene, many cave systems within the basin, which were interconnected previously, would have become completely or periodically isolated, preventing gene flow among populations in close proximity, giving rise to the patterns among sister taxa that we see today. Interestingly, gene flow is more prominent around the karstic northwestern regions of the Guangxi plains, especially in Clades B and C. Therefore, we can assume that the deeper caves and the subterranean river systems associated with the Guangxi karst region (Zhao and Zhang, 2009) were periodically interconnected, allowing limited gene flow. The deeper phylogenetic relationship of *S. tianlinensis* and the species in Clade B, suggests that they diverged during the Miocene. Furthermore, introgression is not apparent among the species from the hilly terrains of Yunnan plateau, suggesting that they inhabit isolated subterranean systems in these hills.

5. CONCLUDING REMARKS

In summary, our study showed that: 1) instead of the 4 major clades of *Sinocyclocheilus* previously recognized by phylogenetic relationships, the genus can be now categorized into 6 major clades; 2) the MRCA of *Sinocyclocheilus* appeared around 10.5 Mya coinciding with cave formation due to upliftment and dry conditions associated with the aridification of China during the late Miocene and the Pliocene; 3) the BFD* analyses support the hypothesis that the two cryptic species found in this study (*S.* cf. *longibarbatus* and *S.* cf. *guanyangensis*) and the morphologically similar *S. tianeensis* and *S. furcodorsalis* are distinct species. We further draw attention to the fact that the diversity of the *Sinocyclocheilus* species may be substantially underestimated due to their cryptic nature. Future studies using novel genomic techniques (such as RADseq or whole genome sequencing) could potentially unravel the true diversity of this remarkable radiation.

Abbreviations

RADseq: Restriction site-associated DNA sequencing; mt-DNA: Mitochondrial DNA; BFD: Bayes factor delimitation; Mya: Million years ago; ILS: Incomplete lineage sorting; PCA: Principal component analysis; SNP: Single-nucleotide polymorphism; MSC: Multispecies coalescent; MLE: Marginal likelihood estimate; ML: Maximum Likelihood

Acknowledgements

We thank Rohan Pethiyagoda (Australian Museum) and Mario van Gastel for suggestions to improve the paper. We thank the following for assistance in the field: Shipeng Zhou, Bing Chen, Dan Sun, Jayampathi Herath and Amrapali Rajput.

Authors' contributions

MM, MMV, MRP, TRM, YWL, conceptualized the research and designed the methodology. YWL, CHF, MTR, MM, GE, JY conducted fieldwork and curated the data. TRM, MMV, YWL, MRP, GE, carried out formal analysis. TRM, MMV, MM, MRP, GE, YWL, wrote the original draft. MM, MMV, MRP, and JY supervised MTR & YWL. MM and JY acquired funding. TMR, GE and YWL made figures. All authors reviewed and edited the draft. All authors read and approved the final manuscript.

Funding

Funding for this study is provided by (1) Guangxi University Startup Funding to MM for fieldwork, lab work, analyses and supporting TRM, YWL, CHF (2) National Natural Science Foundation of China (#31860600) to JY for fieldwork (3) Guangxi Natural Science Foundation (#2017GXNSFFA198010) to JY for research work. These funding bodies played no role in the design of the study and collection, analysis, and interpretation of data or in the writing of the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included as supplementary information and the genetic data (Genbank) can be accessed upon acceptance of the paper. Please see the materials and methods section for the SRA project name for RAD data.

Ethics approval and permission

Methods of sampling approval by the Ethics Committee of Guangxi University. Field sampling approval through the Guangxi Provincial Government.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

6. REFERENCES

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A.J.P.o., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. 3, e3376.

Borowsky, R.J.B.S.F.E.S.P., 2010. The evolutionary genetics of Cave fishes: convergence, adaptation and pleiotropy. 141-168.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J.J.P.c.b., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. 10, e1003537.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A.J.M.b., evolution, 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. 29, 1917-1932.

Cariou, M., Duret, L., Charlat, S.J.E., evolution, 2013. Is RAD‐seq suitable for phylogenetic inference? An in silico assessment and optimization. 3, 846-852.

Chen, S., Zhang, R., Feng, J., Xiao, H., Li, W., Zan, R., Zhang, Y.J.J.o.F.B., 2009. Exploring factors shaping population genetic structure of the freshwater fish Sinocyclocheilus grahami (Teleostei, Cyprinidae). 74, 1774-1786.

Chen, Y.Y., Li, R., Li, C.Q., Li, W.X., Yang, H.F., Xiao, H., Chen, S.Y., 2018. Testing the validity of two putative sympatric species from Sinocyclocheilus (Cypriniformes: Cyprinidae) based on mitochondrial cytochrome b sequences. Zootaxa 4476, 130-140.

Chifman, J., Kubatko, L.J.B., 2014. Quartet inference from SNP data under the coalescent model. 30, 3317-3324.

Eaton, D.A., Overcast, I.J.B., 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. 36, 2592-2594.

He, Y., Chen, X.-Y., Xiao, T.-Q., Yang, J.-X., 2013. Three-dimensional morphology of the Sinocyclocheilus hyalinus (Cypriniformes: Cyprinidae) horn based on synchrotron X-ray microtomography.

Herrera, S., Shank, T.M.J.M.P., Evolution, 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. 100, 70-79.

Hewitt, G.J.N., 2000. The genetic legacy of the Quaternary ice ages. 405, 907-913.

Hutchison, V.H.J.E.M., 1958. The distribution and ecology of the cave salamander, Eurycea lucifuga. 28, 2-20.

Jeffery, W.R., 2019. Astyanax mexicanus: A vertebrate model for evolution, adaptation, and development in caves. Encyclopedia of Caves. Elsevier, pp. 85-93.

Jeffery, W.R., Strickler, A.G., Trajano, E., Bichuette, M., Kapoor, B.J.B.o.S.F.E.S.P., 2010. Development as an evolutionary process in Astyanax cavefish. 141-168.

Jiang, W.-S., Li, J., Lei, X.-Z., Wen, Z.-R., Han, Y.-Z., Yang, J.-X., Chang, J.-B.J.Z.r., 2019. Sinocyclocheilus sanxiaensis, a new blind fish from the Three Gorges of Yangtze River provides insights into speciation of Chinese cavefish. 40, 552.

Kim, D., Bauer, B.H., Near, T.J.J.S.B., 2021. Introgression and Species Delimitation in the Longear Sunfish Lepomis megalotis (Teleostei: Percomorpha: Centrarchidae).

Klaus, S., Mendoza, J.C., Liew, J.H., Plath, M., Meier, R., Yeo, D.C.J.B.l., 2013. Rapid evolution of troglomorphic characters suggests selection rather than neutral mutation as a driver of eye reduction in cave crabs. 9, 20121098.

Leaché, A.D., Banbury, B.L., Felsenstein, J., De Oca, A.N.-M., Stamatakis, A.J.S.b., 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. 64, 1032-1047.

Leaché, A.D., Bouckaert, R.R., 2018. Species trees and species delimitation with SNAPP: a tutorial and worked example. Workshop on Population and Speciation Genomics, Český Krumlov.

Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R.J.S.b., 2014. Species delimitation using genome-wide SNP data. 63, 534-542.

Leal-Zanchet, A.M., de Souza, S.T., Ferreira, R.L., 2014. A new genus and species for the first recorded cave-dwelling Cavernicola (Platyhelminthes) from South America. Zookeys, 1-15.

Li, C., Chen, H., Zhao, Y., Chen, S., Xiao, H.J.E., evolution, 2020. Comparative transcriptomics reveals the molecular genetic basis of pigmentation loss in Sinocyclocheilus cavefishes. 10, 14256-14271.

Li, J., Fang, X.J.C.S.B., 1999. Uplift of the Tibetan Plateau and environmental changes. 44, 2117-2124.

Li, X., Guo, B.J.P.o.t.R.S.B., 2020. Substantially adaptive potential in polyploid cyprinid fishes: evidence from biogeographic, phylogenetic and genomic studies. 287, 20193008.

Li, Z., Guo, B., Li, J., He, S., Chen, Y.J.C.S.B., 2008. Bayesian mixed models and divergence time estimation of Chinese cavefishes (Cyprinidae: Sinocyclocheilus). 53, 2342-2352.

Li, Z., He, S.J.H., 2009. Relaxed purifying selection of rhodopsin gene within a Chinese endemic cavefish genus Sinocyclocheilus (Pisces: Cypriniformes). 624, 139-149.

Liang, X.-F., Cao, L., Zhang, C.-g.J.E.b.o.f., 2011. Molecular phylogeny of the Sinocyclocheilus (Cypriniformes: Cyprinidae) fishes in northwest part of Guangxi, China. 92, 371-379.

Ma, L., Zhao, Y., Yang, J.-x., 2019. Cavefish of China. Encyclopedia of caves. Elsevier, pp. 237-254.

Ma, Z., Herzog, H., Jiang, Y., Zhao, Y., Zhang, D.J.I.z., 2020. Exquisite structure of the lateral line system in eyeless cavefish Sinocyclocheilus tianlinensis contrast to eyed Sinocyclocheilus macrophthalmus (Cypriniformes: Cyprinidae). 15, 314-328.

Mao, T.-R., Liu, Y.-W., Meegaskumbura, M., Yang, J., Ellepola, G., Senevirathne, G., Fu, C.-H., Gross, J.B., Pie, M.R.J.B.e., evolution, 2021. Evolution in Sinocyclocheilus cavefish is marked by rate shifts, reversals, and origin of novel traits. 21, 1-14.

Meng, F., Braasch, I., Phillips, J.B., Lin, X., Titus, T., Zhang, C., Postlethwait, J.H.J.M.b., evolution, 2013. Evolution of the eye transcriptome under constant darkness in Sinocyclocheilus cavefish. 30, 1527-1543.

Pickrell, J., Pritchard, J.J.N.P., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. 1-1.

Popovic, I., Matias, A.M.A., Bierne, N., Riginos, C.J.E.A., 2020. Twin introductions by independent invader mussel lineages are both associated with recent admixture with a native congener in Australia. 13, 515-532.

Porter, M.L., Crandall, K.A.J.T.i.E., Evolution, 2003. Lost along the way: the significance of evolution in reverse. 18, 541-547.

Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M.-R., Elmer, K.R., Vences, M., Steinfarz, S.J.M.p., evolution, 2019. Phylogeny and species delimitation of near Eastern Neurergus newts (Salamandridae) based on genome-wide RADseq data analysis. 133, 189-197.

Razkin, O., Sonet, G., Breugelmans, K., Madeira, M.J., Gómez-Moliner, B.J., Backeljau, T.J.M.P., Evolution, 2016. Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex Pyramidula: the power of RADseq data. 101, 267-278.

Romero, A., Zhao, Y., Chen, X., 2009. The hypogean fishes of China. Chinese Fishes. Springer, pp. 211-278.

Shi, Y., Li, J., Li, B.J.A.G.S.-C.E.-. 1999. Uplift of the Qinghai-Xizang (Tibetan) plateau and east Asia environmental change during late Cenozoic. 54, 20-28.

Stamatakis, A.J.B., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. 30, 1312-1313.

Stange, M., Sánchez-Villagra, M.R., Salzburger, W., Matschiner, M.J.S.b., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. 67, 681-699.

Svendsen, J.I., Alexanderson, H., Astakhov, V.I., Demidov, I., Dowdeswell, J.A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., Houmark-Nielsen, M.J.Q.S.R., 2004. Late Quaternary ice sheet history of northern Eurasia. 23, 1229-1271.

Wilgenbusch, J.C., Swofford, D.J.C.p.i.b., 2003. Inferring evolutionary trees with PAUP. 6.4. 1-6.4. 28.

Xiao, H., Chen, S.-y., Liu, Z.-m., Zhang, R.-d., Li, W.-x., Zan, R.-g., Zhang, Y.-p.J.M.P., Evolution, 2005. Molecular phylogeny of Sinocyclocheilus (Cypriniformes: Cyprinidae) inferred from mitochondrial DNA sequences. 36, 67-77.

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S.J.B.b., 2016. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. 14, 1-13.

Yang, N., Li, Y., Liu, Z., Chen, Q., Shen, Y.J.E.B.o.F., 2021. Molecular phylogenetics and evolutionary history of Sinocyclocheilus (Cypriniformes: Cyprinidae) species within Barbinae in China. 1-14.

Zhao, Y., Zhang, C., 2009. Endemic fishes of Sinocyclocheilus (Cypriniformes: Cyprinidae) in China-species diversity, cave adaptation, systematics and zoogeography. Beijing: Science Press.

MAIN TEXT TABLES AND FIGURES

Table 1. Description of the BFD* species delimitation models for the *Sinocyclocheilus* groups, including the model tested, the number of species, the MLE, the BF between that model and the model of current taxonomy, the model rank and the clade to different species models belong. Since the unidentified species are sister groups with *S. guanyangensis* and *S. longibarbatus*, we herein label them as *S.* cf. *guanyangensis* and *S.* cf. *longibarbatus*.

Model	Numbers of Species	MLE	Rank	BF	Clade
Current taxonomy: ((*S. guanyangensis*, *S. huangtianensis*), *S. guilinensis*)	3	-2044.0811	2	-	A
ModelA1: (((*S. guanyangensis*, S. cf. *guanyangensis*), *S. huangtianensis*), *S. guilinensis*)	4	-2043.4052	1	1.3517	A
ModelA2: (*S. huangtianensis*, *S. guilinensis*)	2	-2220.3862	3	-352.6102	A
Current taxonomy: ((*S. longibarbatus*, *S. xunlensis*), (*S. huanjiangensis*, *S. brevis*))	4	-1608.1623	2	-	C
ModelC1: (((*S. longibarbatus*, S. cf. *longibarbatus*), *S. xunlensis*), (*S. huanjiangensis*, *S. brevis*))	5	-1581.8094	1	52.7057	C
ModelC2: ((*S. brevis*, *S. huanjiangensis*), *S. xunlensis*)	3	-1691.0134	3	-165.7023	C
ModelC3: (*S. brevis*, *S. xunlensis*)	2	-1783.1874	4	-350.0501	C
Current taxonomy: (((*S. mashanensis*, *S. brevibarbatus*), *S. altishoulderus*), (*S. tianeensis*, *S. furcodorsalis*))	5	-1564.9927	1	-	B
ModelB1: (((*S. mashanensis*, *S. brevibarbatus*), *S. altishoulderus*), *S. furcodorsalis*)	4	-1572.4045	2	-14.8236	B

ModelB2: ((*S. brevibarbatus*, *S. altishoulderus*), *S. furcodorsalis*)	3	-1596.9541	4	-63.9227	B
ModelB3: (*S. altishoulderus*, *S. furcodorsalis*)	2	-1678.6393	3	-227.2931	B

*MLE = Marginal likelihood estimate
*BF = Bayes factor

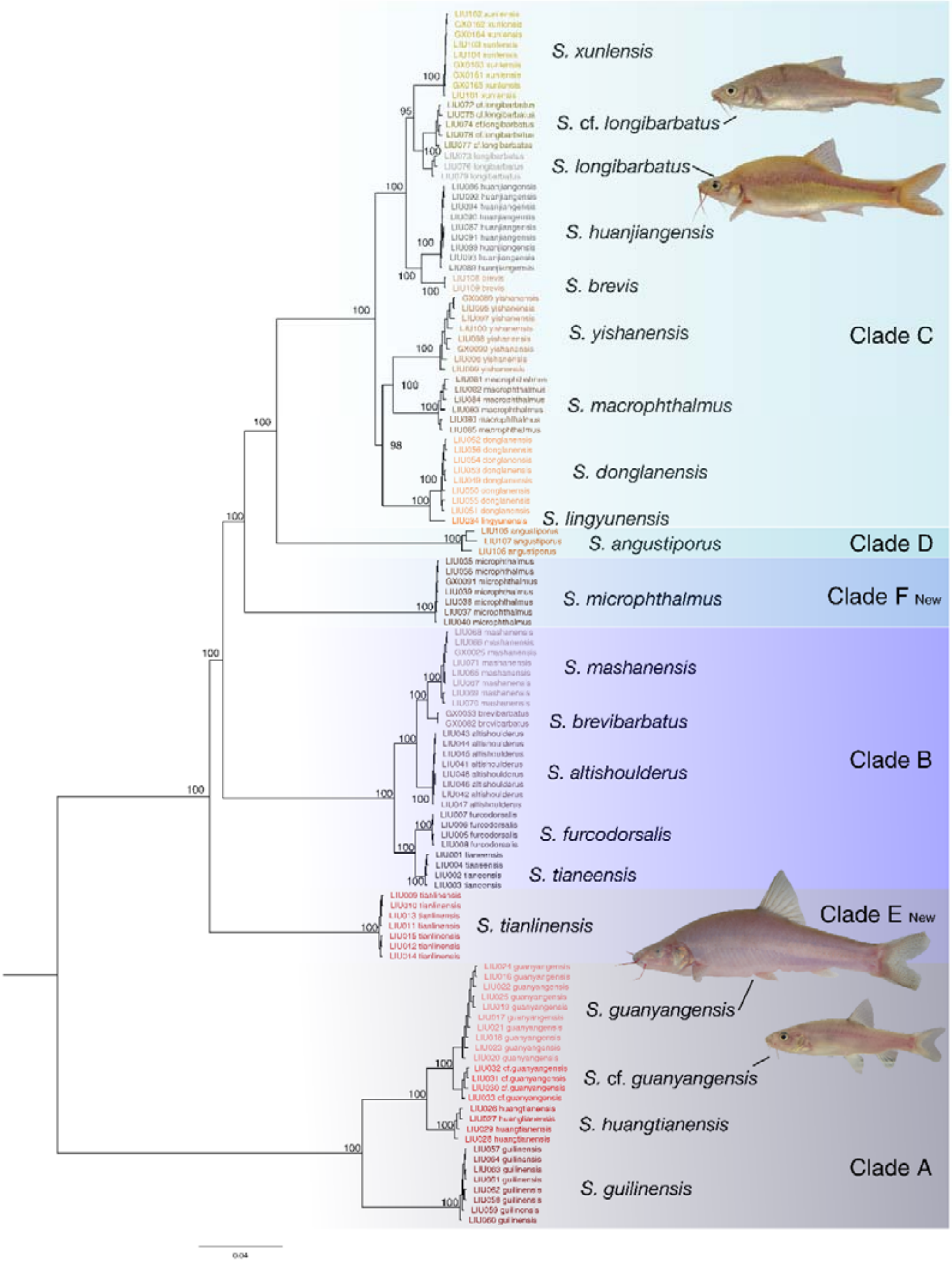

Fig 1. Phylogenomic relationships of the *Sinocyclocheilus* species based on the unpartitioned concatenated maximum likelihood (ML) analysis of a matrix with 646,497 characters using a GTRCAT model of substitution. Bootstrap values are displayed on the nodes.

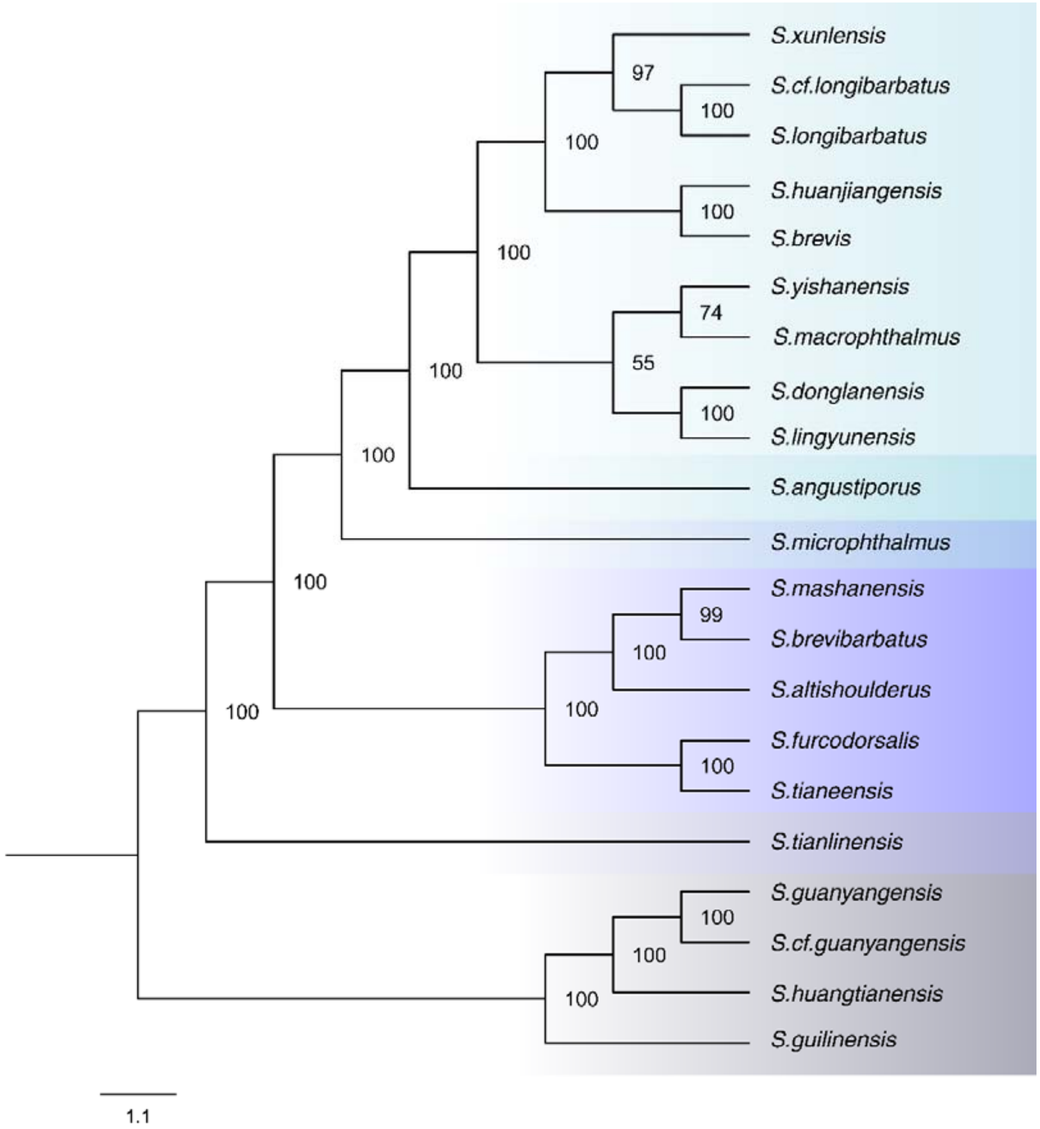

Fig 2. Topology estimated from SVDquartets species tree analysis indicating the phylogenetic relationships of all the 21 *Sinocyclocheilus* species and bloodlines identified in our study based on 4378 unlinked SNPs for 120 individuals.

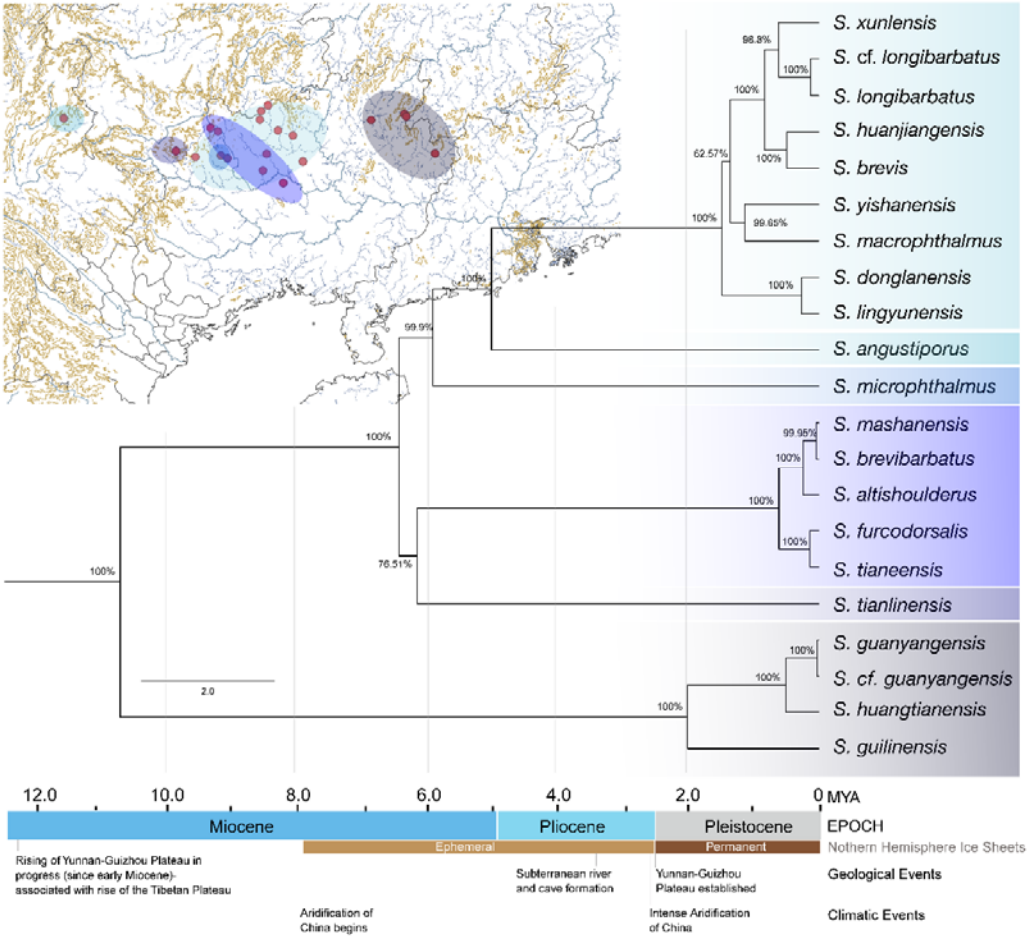

Fig 3. Time-calibrated maximum clade credibility species tree of the all 21 *Sinocyclocheilus* species is inferred by SNAPP. Node support is indicated by the Bayesian posterior probability next to nodes.

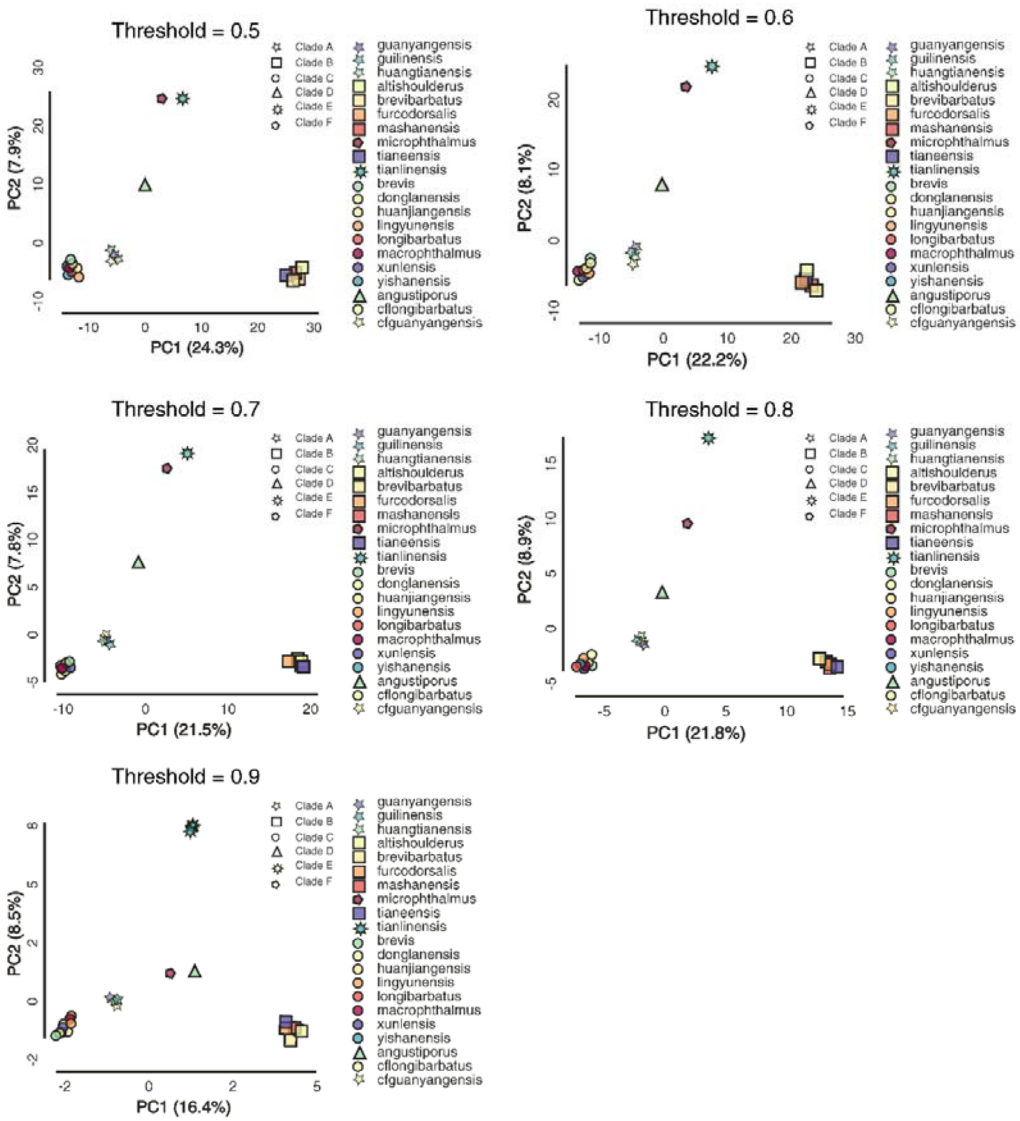

Fig 4. The PCA plots based on the 3459, 2942, 1993, 987, 165 SNP dataset show the population clustering between the 120 samples. The projection of individual samples on the surface is defined by the first two axes of the principal components, the x-axis (PC1 = 24.3%, 22.2%, 21.5%, 21.8%, 16.4%) and the y-axis (PC2 = 7.9%, 8.1%, 7.8%, 8.9%, 8.5%). Different symbols corresponded to the different clades. Different colors corresponded to differentiate the species inside each clade.

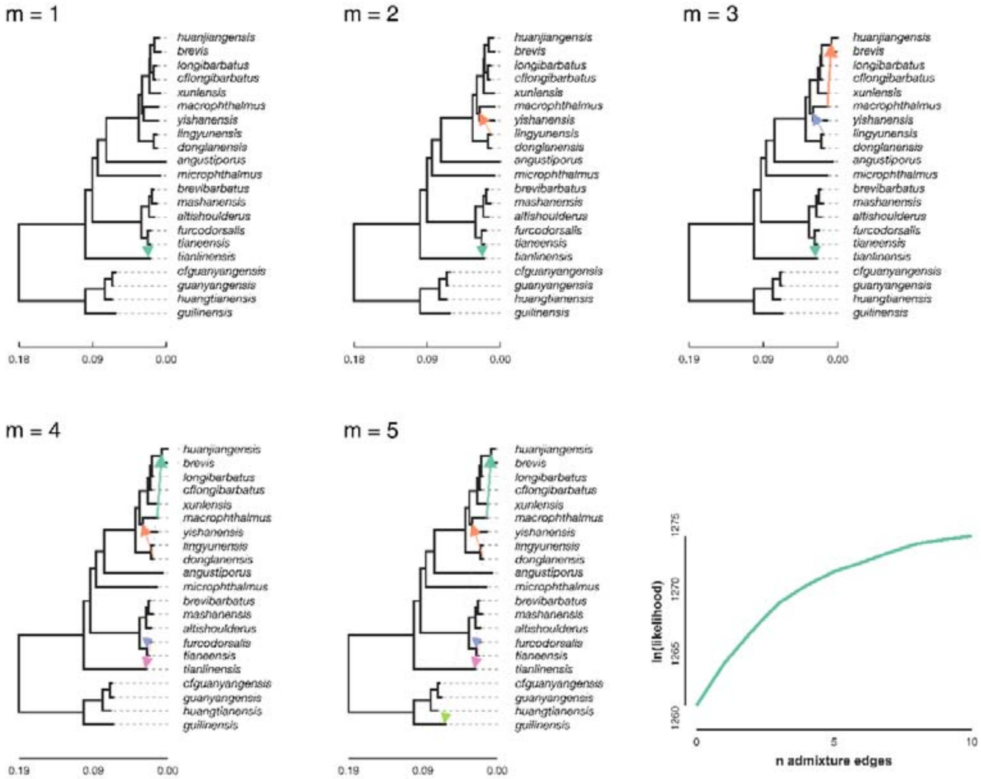

Fig 5. Treemix trees with different admixture edges (m= 1-5) and the plotting of the ln(likelihood) for different values of admixture edges. Arrows indicate when migration occurred in the population tree. The thickness represents the strength of the migration weight.

SUPPORTING INFORMATION

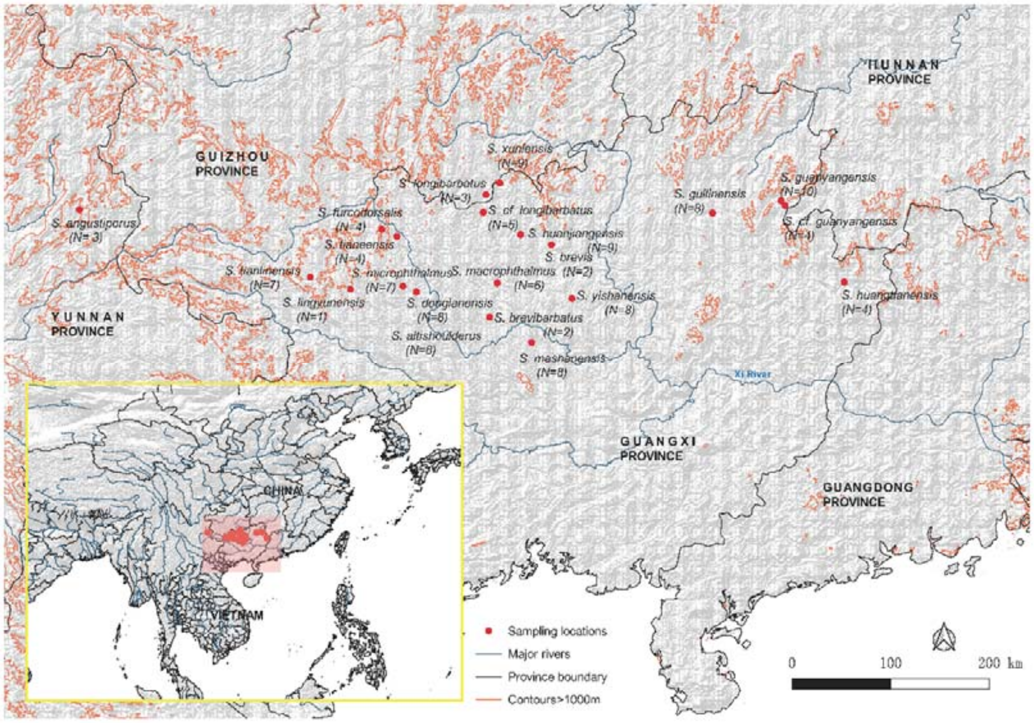

Fig S1. Map of the sampling localities of *Sinocyclocheilus* species. *S.donglanensis* and *S.altishoulderus* are caught in the same cave.

Table S1. Summary of the RAD data

Species	Sample ID	reads_raw	reads_passed_filter	Retained reads (%)	clusters_total_at 90%	clusters_hidepth	hetero_est	error_est	reads_consens	loci_in_assembly_m114
S.mashanensis	GX0025mashanensis	6147831	6142674	0.999161168	620708	192493	0.010226	0.003766	175189	4019
S.brevibarbatus	GX0053brevibarbatus	10574223	10566811	0.99929905	718281	294336	0.009814	0.002564	272278	4334
S.brevibarbatus	GX0082brevibarbatus	11322894	11314123	0.999225375	760656	330506	0.009505	0.002611	306178	4340
S.yishanensis	GX0089yishanensis	11666117	11658167	0.999318539	813824	351254	0.010757	0.002407	324895	4337
S.yishanensis	GX0090yishanensis	10898276	10886321	0.998903038	1086550	337981	0.011189	0.002539	311630	4244
S.microphthalmus	GX0091microphthalmus	12193551	12185058	0.999303484	739666	320505	0.007106	0.003299	298095	4340
S.xunlensis	GX0161xunlensis	14567416	14535854	0.997833384	714585	332393	0.007879	0.001922	310820	4363
S.xunlensis	GX0162xunlensis	8286436	8262346	0.997092839	636925	262650	0.009182	0.002243	243443	4309
S.xunlensis	GX0163xunlensis	10705744	10695277	0.999022301	655543	304495	0.008336	0.002137	283693	4356
S.xunlensis	GX0164xunlensis	10123039	10112284	0.998937572	702603	289488	0.008326	0.002259	269436	4361
S.xunlensis	GX0165xunlensis	7784571	7775855	0.998880349	599966	255365	0.00899	0.002364	236322	4279
S.tianeensis	LIU001tianeensis	6200451	6197355	0.999500681	611306	183928	0.010657	0.003789	167325	3995
S.tianeensis	LIU002tianeensis	7039053	7035295	0.999466121	652330	211343	0.01034	0.00397	192887	4137
S.tianeensis	LIU003tianeensis	11888599	11881038	0.999364013	742560	293098	0.008929	0.003864	271517	4324
S.tianeensis	LIU004tianeensis	14921317	14911227	0.999323786	826400	324984	0.008396	0.004097	301144	4336
S.furcodorsalis	LIU005furcodorsalis	10769061	10760550	0.99920968	694537	281720	0.008468	0.003948	260984	4312
S.furcodorsalis	LIU006furcodorsalis	10372715	10362650	0.999029666	667800	281437	0.009014	0.003599	259945	4230
S.furcodorsalis	LIU007furcodorsalis	9634251	9627253	0.999273633	686353	248630	0.008375	0.003833	230088	4308
S.furcodorsalis	LIU008furcodorsalis	9696241	9690370	0.999394508	681881	265864	0.00886	0.00373	245785	4319
S.tianlinensis	LIU009tianlinensis	8995569	8987159	0.999065095	650273	250974	0.010398	0.002293	230813	4128
S.tianlinensis	LIU010tianlinensis	11999075	11988071	0.999082929	672240	305230	0.009288	0.002118	283459	4338
S.tianlinensis	LIU011tianlinensis	12705218	12692693	0.999014185	717785	298209	0.008986	0.002306	276261	4319
S.tianlinensis	LIU012tianlinensis	9967212	9958416	0.999117506	663287	278613	0.010455	0.002181	256958	4268
S.tianlinensis	LIU013tianlinensis	9446480	9434427	0.998724075	657929	263438	0.010548	0.002229	243003	4247
S.tianlinensis	LIU014tianlinensis	14062403	14037489	0.998228326	666170	301330	0.009456	0.001876	279731	4139

S.tianlinensis	LIU015tianlinensis	9837696	9828401	0.999055165	643895	275394	0.010283	0.002259	254319	4277
S.guanyangensis	LIU016guanyangensis	5835416	5831140	0.999267233	575893	191385	0.010975	0.002485	175146	4081
S.guanyangensis	LIU017guanyangensis	7425008	7420300	0.999365927	595404	247260	0.010159	0.002249	228375	4307
S.guanyangensis	LIU018guanyangensis	7599328	7594728	0.999394683	596396	244371	0.009842	0.002445	225969	4302
S.guanyangensis	LIU019guanyangensis	5502861	5498020	0.999120276	552348	183387	0.012421	0.002477	166198	3953
S.guanyangensis	LIU020guanyangensis	9618318	9610904	0.999229179	644403	272336	0.009145	0.00284	251941	4292
S.guanyangensis	LIU021guanyangensis	9791336	9784948	0.999347586	663951	277438	0.009687	0.00269	256631	4326
S.guanyangensis	LIU022guanyangensis	9641947	9635302	0.999310824	657512	273533	0.009717	0.002758	253231	4340
S.guanyangensis	LIU023guanyangensis	8690802	8685388	0.999377043	639812	263571	0.009773	0.002659	243131	4318
S.guanyangensis	LIU024guanyangensis	9929201	9920990	0.999173045	726970	277238	0.00917	0.003504	256422	4331
S.guanyangensis	LIU025guanyangensis	8665897	8658838	0.999185428	769636	237526	0.01096	0.003681	216446	4236
S.guanyangensis	LIU026huangtianensis	7999484	7993211	0.999215824	760252	238732	0.011484	0.003646	217323	4190
S.guanyangensis	LIU027huangtianensis	11407539	11398034	0.999166779	829908	291284	0.010385	0.003511	268354	4291
S.guanyangensis	LIU028huangtianensis	11051312	11042598	0.999211496	779445	287736	0.0102	0.003384	265541	4297
S.guanyangensis	LIU029huangtianensis	7431568	7423445	0.99890696	690005	229699	0.011089	0.0035	210050	4163
S.cfguanyangensis	LIU030guanyangensis	10367199	10354231	0.998749132	706623	284698	0.009284	0.003026	263659	4309
S.cfguanyangensis	LIU031guanyangensis	15698789	15686607	0.999224017	859130	332964	0.009374	0.0033	309120	4321
S.cfguanyangensis	LIU032guanyangensis	11610896	11601005	0.999148128	739162	300310	0.009589	0.00284	278115	4302
S.cfguanyangensis	LIU033guanyangensis	13628426	13613358	0.99889437	766450	329302	0.009591	0.002964	305771	4333
S.lingyunensis	LIU034lingyunensis	10200248	10193547	0.999343055	705087	275869	0.008302	0.003357	255620	4276
S.microphthalmus	LIU035microphthalmus	9646294	9640033	0.999350942	666565	247166	0.007995	0.0033	228746	4260
S.microphthalmus	LIU036microphthalmus	11684750	11676797	0.999319369	687515	278963	0.007914	0.003059	259306	4317
S.microphthalmus	LIU037microphthalmus	7203537	7197587	0.999174017	626951	214213	0.0087	0.003222	197446	4165
S.microphthalmus	LIU038microphthalmus	10711256	10703957	0.999318567	761746	274957	0.008384	0.003455	254268	4299
S.microphthalmus	LIU039microphthalmus	13168573	13157998	0.999196952	787637	315043	0.007626	0.003866	293031	4342
S.microphthalmus	LIU040microphthalmus	10504013	10495614	0.999200401	729223	280416	0.008512	0.00306	259710	4303
S.altishoulderus	LIU041altishoulderus	9115942	9110083	0.99935728	712724	242381	0.009619	0.003367	223200	4279
S.altishoulderus	LIU042altishoulderus	9295671	9290264	0.999418331	680978	253956	0.009416	0.003366	234614	4307
S.altishoulderus	LIU043altishoulderus	3686672	3684018	0.99928011	530351	91735	0.013846	0.003929	80739	2676

S.altishoulderus	LIU044altishoulderus	8829672	8823233	0.999270754	710418	238946	0.009373	0.003716	220079	4314
S.altishoulderus	LIU045altishoulderus	6253630	6248873	0.999239322	626661	184394	0.010154	0.00361	168412	4110
S.altishoulderus	LIU046altishoulderus	10379052	10371057	0.999229698	726338	267972	0.009443	0.003428	247545	4300
S.altishoulderus	LIU047altishoulderus	11990325	11981108	0.999231297	761782	280146	0.009103	0.003454	259243	4301
S.altishoulderus	LIU048altishoulderus	9007952	9000403	0.999161963	661993	235686	0.009187	0.00345	217407	4309
S.donglanensis	LIU049donglanensis	7863731	7858427	0.999325511	650142	217966	0.009355	0.00341	200740	4269
S.donglanensis	LIU050donglanensis	6012065	6008402	0.999390725	595334	182885	0.009644	0.003311	167463	4101
S.donglanensis	LIU051donglanensis	7270422	7264479	0.999182578	645058	216115	0.009511	0.003288	198658	4257
S.donglanensis	LIU052donglanensis	5189844	5185630	0.99918803	589539	158550	0.010208	0.0034	144164	3888
S.donglanensis	LIU053donglanensis	7049142	7043507	0.999200612	639506	202634	0.009277	0.003351	186106	4245
S.donglanensis	LIU054donglanensis	7100565	7092862	0.998915157	641196	208822	0.009545	0.003225	192016	4249
S.donglanensis	LIU055donglanensis	11642270	11633291	0.999228759	778733	273183	0.008489	0.00331	252643	4319
S.donglanensis	LIU056donglanensis	9902488	9896300	0.999375107	710436	261784	0.008648	0.003204	242095	4317
S.guilinensis	LIU057guilinensis	9332971	9324079	0.999047249	709154	238640	0.009599	0.003527	220326	4267
S.guilinensis	LIU058guilinensis	8565978	8558576	0.999135884	684331	234549	0.009204	0.003282	216566	4241
S.guilinensis	LIU059guilinensis	10775383	10764841	0.999021659	789590	262483	0.009461	0.003997	242317	4297
S.guilinensis	LIU060guilinensis	9524201	9514228	0.998952878	840547	246917	0.010395	0.004384	224668	4210
S.guilinensis	LIU061guilinensis	12009748	11999083	0.999111971	831139	287828	0.008884	0.003979	265617	4309
S.guilinensis	LIU062guilinensis	11881410	11869513	0.998998688	930238	285733	0.010819	0.00418	260322	4219
S.guilinensis	LIU063guilinensis	5573070	5567744	0.999044333	716929	163633	0.013349	0.004489	144726	3700
S.guilinensis	LIU064guilinensis	6429539	6421534	0.998754965	646217	196636	0.010554	0.003891	179618	4034
S.mashanensis	LIU065mashanensis	12773768	12759456	0.998879579	771966	302170	0.008833	0.003317	280230	4325
S.mashanensis	LIU066mashanensis	18924069	18912035	0.99936409	895580	348784	0.007795	0.003601	324401	4325
S.mashanensis	LIU067mashanensis	8793721	8787492	0.999291654	677670	255773	0.009239	0.003624	235973	4286
S.mashanensis	LIU068mashanensis	9602601	9593898	0.999093683	689748	255671	0.008455	0.003662	236501	4328
S.mashanensis	LIU069mashanensis	18354043	18341140	0.999296994	898018	367559	0.008771	0.003904	341854	4349
S.mashanensis	LIU070mashanensis	9315134	9308690	0.999308223	676300	251238	0.008786	0.003883	231899	4325
S.mashanensis	LIU071mashanensis	11296508	11288842	0.999321383	716126	281199	0.008727	0.003443	260513	4340
S.cflongibarbatus	LIU072longibarbatus	12556306	12546046	0.999182881	819994	295806	0.009112	0.003407	273901	4323

S.longibarbatus	LIU073longibarbatus	11256884	11246465	0.999074433	779530	278082	0.00846	0.004172	257838	4334
S.cflongibarbatus	LIU074longibarbatus	8672867	8665361	0.999134542	684535	244397	0.009918	0.003615	224987	4314
S.cflongibarbatus	LIU075longibarbatus	10392345	10383838	0.999181417	757547	285335	0.009168	0.003603	263834	4317
S.longibarbatus	LIU076longibarbatus	9349351	9342131	0.999227754	729563	268365	0.00913	0.003531	247476	4312
S.cflongibarbatus	LIU077longibarbatus	8170222	8163676	0.999198798	689396	232697	0.010145	0.003514	214146	4294
S.cflongibarbatus	LIU078longibarbatus	8740967	8734259	0.999232579	666300	240461	0.010657	0.002464	222145	4324
S.longibarbatus	LIU079longibarbatus	8238575	8231454	0.999135651	684689	227551	0.010435	0.00277	209521	4305
S.macrophthalmus	LIU080macrophthalmus	8766561	8758914	0.999127708	699118	245249	0.010893	0.002666	225910	4305
S.macrophthalmus	LIU081macrophthalmus	10822267	10810591	0.998921113	778952	278653	0.010358	0.002571	257592	4315
S.macrophthalmus	LIU082macrophthalmus	11554037	11542467	0.998998618	695695	289584	0.009913	0.002365	268091	4325
S.macrophthalmus	LIU083macrophthalmus	10235217	10224387	0.998941889	653383	272724	0.010031	0.002186	252622	4335
S.macrophthalmus	LIU084macrophthalmus	10150601	10141912	0.999143992	675414	273280	0.010294	0.00258	252910	4326
S.macrophthalmus	LIU085macrophthalmus	8387902	8381373	0.999221617	627774	231344	0.010422	0.0025	213273	4295
S.huanjiangensis	LIU086huanjiangensis	9860668	9853911	0.999314752	638229	269510	0.008114	0.002447	250258	4355
S.huanjiangensis	LIU087huanjiangensis	9216530	9210709	0.999368417	651220	253423	0.008296	0.002563	235163	4333
S.huanjiangensis	LIU088huanjiangensis	10088282	10083081	0.999484451	635342	278161	0.008295	0.002575	258408	4324
S.huanjiangensis	LIU089huanjiangensis	8970782	8963190	0.999153697	639357	250469	0.00889	0.002524	231848	4322
S.huanjiangensis	LIU090huanjiangensis	11690809	11680339	0.999104425	778832	287701	0.008311	0.003494	266758	4322
S.huanjiangensis	LIU091huanjiangensis	9812008	9805062	0.999292092	698569	264968	0.008369	0.003467	245471	4337
S.huanjiangensis	LIU092huanjiangensis	10110052	10101594	0.999163407	729917	263698	0.007999	0.003659	244153	4337
S.huanjiangensis	LIU093huanjiangensis	9490144	9483135	0.999261444	704181	277437	0.008368	0.003408	257015	4332
S.huanjiangensis	LIU094huanjiangensis	10331954	10324059	0.999235866	678557	266485	0.008769	0.002702	247105	4349
S.yishanensis	LIU095yishanensis	11078825	11070738	0.999270049	689417	278309	0.01088	0.002671	257159	4323
S.yishanensis	LIU096yishanensis	9209644	9203319	0.99931322	668856	258976	0.011428	0.002687	238602	4297
S.yishanensis	LIU097yishanensis	11948928	11939669	0.999225119	710819	280410	0.010595	0.00249	259199	4321
S.yishanensis	LIU098yishanensis	9717902	9711389	0.999329794	627159	253618	0.010658	0.00225	234931	4316
S.yishanensis	LIU099yishanensis	8253934	8245811	0.999015863	640514	237456	0.011343	0.002597	218144	4295
S.yishanensis	LIU100yishanensis	9472530	9461442	0.998829457	626280	253619	0.010567	0.002346	234567	4315
S.xunlensis	LIU101xunlensis	8717881	8708887	0.998968327	648178	239426	0.008968	0.002481	221900	4334

S.xunlensis	LIU102xunlensis	9176341	9169184	0.999220059	658452	256582	0.008473	0.00257	237853	4336
S.xunlensis	LIU103xunlensis	11154384	11142474	0.998932258	773286	301312	0.009184	0.002294	279735	4353
S.xunlensis	LIU104xunlensis	18736734	18717682	0.998983174	1091524	395736	0.007878	0.002228	369084	4356
S.angustiporus	LIU105angustiporus	5236660	5232391	0.999184786	605103	151880	0.016385	0.002671	135907	3236
S.angustiporus	LIU106angustiporus	6884510	6879551	0.999279687	684665	205307	0.016159	0.002812	185042	3688
S.angustiporus	LIU107angustiporus	6737620	6731762	0.999130554	663549	200421	0.014345	0.003028	181613	3597
S.brevis	LIU108brevis	6193310	6188183	0.999172171	599403	194962	0.010927	0.002586	178136	4078
S.brevis	LIU109brevis	9907760	9900681	0.99928551	664376	259661	0.009382	0.002423	240663	4294
Summary		**9832110.98**	**9823582.908**	**0.99913263**	**703204.1167**	**261033.9833**	**0.009693742**	**0.0030865**	**240994.6917**	**4241.525**

Note: values in the last line are averages. Clusters_total: Clusters that passed filtering for 10x minimum coverage. Loci_in_assembly: Loci retained after passing coverage and paralog filters.

Table S2. Sequence information with different parameters

Matrix	**Unlinked SNPs**	**Consensus sequences (bp)**	**VAR**	**PIS**	**Missing (%)**	**PIS/VAR(%)**
Sino_m96	23,869	3,534,842	380,519	336,928	13.46	88.54
Sino_m102	17,187	2,544,628	272,838	242,335	10.89	88.85
Sino_m108	10,549	1,560,414	166,407	148,253	8.21	89.09
Sino_m114	4,378	646,497	67,983	61,023	5.25	89.76
Sino_m120	151	22,274	2,199	1,989	2.29	90.45

Note: Sequence information in the RAD data matrices (n=120) generated with different parameters of minimum number of individuals per locus (m) values. The data matrices shown in bold type were used for analyses.