Molecular Phylogenetics and Evolution 111 (2017) 44–55

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences

Shuli Song, Jinliang Zhao *, Chenhong Li *

Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Ministry of Education, Shanghai 201306, China

ARTICLE INFO

Article history:
Received 25 August 2016
Revised 2 March 2017
Accepted 16 March 2017
Available online 18 March 2017

Keywords:
Sinipercidae
Gene capture
SNPs
BFD*
Historical biogeography
Divergence time

ABSTRACT

The sinipercids are freshwater fishes endemic to East Asia, mainly in China. Phylogenetic studies on the sinipercids have made great progress in the last decades, but interspecific relationships and evolutionary history of the sinipercids remain unresolved. Lack of distinctive morphological characters leads to problems in validating of some species, such as *Siniperca loona*. Moreover, genetic data are needed to delimitate species pairs with explicit hypothesis testing, such as in *S. chuatsi* vs. *S. kneri* and *Coreoperca whiteheadi* vs. *C. liui*. Here we reconstructed phylogeny of the sinipercids with an unprecedented scale of data, 16,943 loci of single-copy coding sequence data from nine sinipercid species, eight putative sister taxa and two outgroups. Targeted sequences were collected using gene enrichment and Illumina sequencing, yielding thousands of protein coding sequences and single nucleotide polymorphisms (SNPs) data. Maximum likelihood and coalescent species tree analyses resulted in identical and highly supported trees. We confirmed that the centrarchids are sister to the sinipercids. A monophyletic Sinipercidae with two genera, *Siniperca* and *Coreoperca* was also supported. Different from most previous studies, *S. scherzeri* was found as the most basal taxon to other species of *Siniperca*, which consists of two clades: a clade having *S. roulei* sister to *S. chuatsi* and *S. kneri*, and a clade consisting *S. loona* sister to *S. obscura* and *S. undulata*. We found that both *S. loona* and *C. liui* are valid species using Bayes factor delimitation (BFD*) based on SNPs data. Species delimitation also provided decisive support for *S. chuatsi* and *S. kneri* being two distinct species. We calibrated a chronogram of the sinipercids based on 100 loci and three fossil calibration points using BEAST, and reconstructed ancestral ranges of the sinipercids using Lagrange Analysis (DEC model) and Statistical Dispersal-Vicariance Analysis (S-DIVA) implemented in RASP. Divergence time estimates and ancestral habitat reconstruction suggested a wide-ranging distribution of the common ancestor of the sinipercids in southern China at 53.1 million years ago (CI: 30.4–85.8 Ma). The calibrated time tree is consistent with historical climate changes and geological events that might have shaped the current distribution of the sinipercids.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Sinipercidae (Perciformes) is a group of freshwater fish distributed in China, Korea, Japan, Russia, and Vietnam with most of its species found only in China. The number of valid species in sinipercids is about 9–12 according to different authors (Li, 1991; Liu and Chen, 1994; Nelson, 2006; Zhou et al., 1988). Substantial progress has been made in resolving phylogenetic relationships among the sinipercids. Traditional classification based on morphology agrees on a monophyletic sinipercids (Liu, 1997; Liu and Chen, 1994), which has been erected as a distinct family (Nelson, 2006; Roberts, 1993). Molecular data also generally support the monophyly of the sinipercids (Li et al., 2010; Near et al., 2012; Zhao et al., 2008, 2006a, 2006b), except that Chen et al. (2007) found that the sinipercids were paraphyletic based on cytochrome b (cytb) sequence.

Morphological studies have competing hypotheses about the sister group of the sinipercids, such as the Serranidae (Jordan, 1923; Nichols, 1943; Zhou et al., 1988) and the Percichthyidae (Gosline, 1966), but consistent results have not been reached (e.g., (Chang, 1988; Johnson, 1984; Liu, 1997; Liu and Chen, 1994; McCully, 1962; Nelson, 1994; Waldman, 1986; Zhou et al., 1988). As for molecular studies, Zhao et al. (2005) found a clade consisting of *Perca* and *Pristigenys* as the sister taxon of the sinipercids. Smith and Craig (2007) supported a clade of the percichthyids

* Corresponding authors.
E-mail addresses: jlzhao@shou.edu.cn (J. Zhao), chli@shou.edu.cn (C. Li).

http://dx.doi.org/10.1016/j.ympev.2017.03.014
1055-7903/© 2017 Elsevier Inc. All rights reserved.

and the centrarchids as the sister taxon of the sinipercids. More recent studies using molecular data recovered a sister-group relationship between the sinipercids and the centrarchids (Chen et al., 2014; Li et al., 2010; Near et al., 2012).

The sinipercids were classified into three genera, *Coreoperca*, *Siniperca* and *Coreosiniperca*, each containing three, seven and one species respectively (Zhou et al., 1988; Zhu, 1985). A cladistic analysis based on 34 morphological characters grouped *Coreosiniperca roulei* with species of *Siniperca* (Liu and Chen, 1994), which was also supported by molecular data (Chen et al., 2010; Li et al., 2010; Zhao et al., 2006a). Although previous molecular studies greatly improved our understanding of the interrelationships among major clades of the Sinipercidae, interspecific relationships among the sinipercids still remain unresolved. Several different molecular phylogenetic analyses aimed at resolving interrelationship among species produced controversial results (Chen et al., 2007, 2010; Li et al., 2010; Zhao et al., 2008, 2006a, 2006b). For example, mitochondrial loci, nuclear loci or combined data supported different relationships among the sinipercids (Fig. 1; (Chen et al., 2010; Li et al., 2010; Zhao et al., 2006a, 2006b). Among those studies, the only congruent result was a confirmed sister relationship between *S. chuatsi* and *S. kneri*, whereas the other relationships were either inconsistent or having low support (Fig. 1).

More important problem challenging systematics of the sinipercids is that the species limits are still unclear. For example, *S. loona* is often considered as a synonym of *S. obscura* (Zhou et al., 1988), but Liu and Chen (1994) argued that *S. loona* should be a valid species based on a couple of morphological characters and its allopatric distribution from *S. obscura* (Fig. S1). Cryptic species was also found in *Coreoperca*. The holotype of *Coreoperca whiteheadi* Boulenger, 1900 was found in the Nandujiang River, Hainan, Chinese southernmost island province. Cao et al. (2013) described a new species, *Coreoperca liui*, on the basis of 19 specimens collected from the Qiantangjiang River basin in Zhejiang province, southeastern China (Fig. S2). The newly described species was similar to *C. whiteheadi* morphologically, only distinguishable in a couple of morphometric characters, but no genetic data were collected to verify the distinctness of the new species. Finally, *S. chuatsi* and *S. kneri* are partially sympatric (Fig. S3), and they have obviously morphological differences in characters, such as number of plyoric-caecum, ratio between eye length and head length (Zhou et al., 1988), but current molecular data (mitochondrial control region) could not separate them as different species (Zhao et al., 2008).

The sinipercids are endemic to East Asia, widely distributed in most Chinese river drainages, so it could represent an interesting group for biogeographic studies on Chinese freshwater fish fauna. Delimiting species boundary and resolving the species phylogeny of the sinipercids are the first step to understand the origin of the sinipercids, the impact of past geological events on their evolutionary patterns, and the evolution of Chinese freshwater fish fauna in general.

A great number of independent loci could result in better-resolved phylogenies (Corl and Ellegren, 2013; Kimball and Braun, 2014). In this study, we applied a targeted gene capture method (Li et al., 2013) to generate a large amount of protein coding sequences and single nucleotide polymorphisms (SNPs) data from eleven species (33 individuals) of the sinipercids, as well as 10 outgroup taxa. Our objectives are: (1) to delimit species status between *C. liui* and *C. whiteheadi*, *S. obscura* and *S. loona*, and *S. chuatsi* and *S. kneri* using multiple independent nuclear loci; (2) to resolve the undetermined interspecific relationships within the sinipercids; (3) and to reconstruct the evolution history of the sinipercids using fossil calibration, with regard to species divergence time and ancestral area reconstruction. Our work may provide a backbone for a better understanding of evolutionary history, biogeography and conservation of the sinipercids.

2. Materials and methods

2.1. Taxon sampling and DNA extraction

Thirty-three individuals from nine species of the sinipercids were sampled and each species was represented by two to ten individuals. The sampling covers all species of the family except for *S. fortis*, *S. liuzhouensis* and *S. robusta* due to availability (Table 1). Ten outgroup species were examined, including *Micropterus salmoides*, *Pomoxis nigromaculatus* (Centrarchidae), *Lateolabrax*

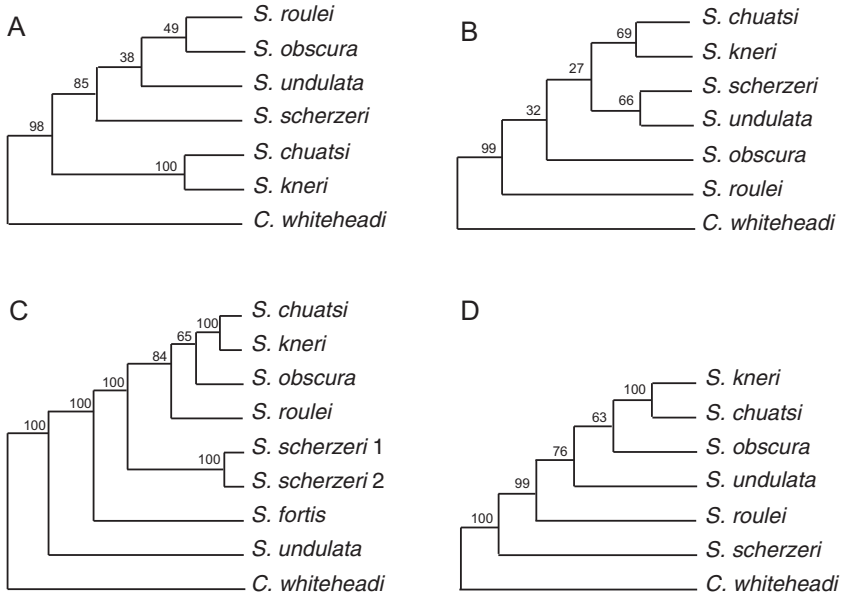

Fig. 1. Different hypotheses about the interrelationship among the sinipercids proposed by previous studies: (A) MP tree based on mitochondrial DNA control region (Zhao et al., 2006b); (B) MP tree obtained from cytochrome *b* sequence (Zhao et al., 2006a); (C) ML phylogeny inferred from combined data including mtDNA, nuclear loci, and viperin (Chen et al., 2010); (D) species tree based on 11 nuclear loci (Li et al., 2010).

Table 1
Taxa sampling, summary of the sequencing results of the 33 sinipercids, eight putative sister taxa and two outgroups.

Voucher number	Species name	Collection location and drainage system	Abbreviation	Collection dates	No. raw reads[a]	No. filtered reads[b]	No. captured loci[c]
CL938_1	*S. chuatsi*	Lake Dongting, Hunan; Yangtze River	YZ 938_1	Jun 30, 2005	33561432	8363230	13442
CL942_1	*S. chuatsi*	Lake Boyang, Jiangxi; Yangtze River	YZ 942_1	Jun 29, 2005	15206272	3787898	12569
CL943_1	*S. chuatsi*	Chizhou, Anhui; Yangtze River	YZ 943_1	Jun 29, 2005	14501920	3611962	12331
CL951_1	*S. chuatsi*	Shunchang, Fujian; Minjiang River	MJ 951-1	Aug 2005	3633648	901364	11381
CL955_1	*S. chuatsi*	Chizhou, Anhui; Yangtze River	YZ 955_1	Jun 6, 2006	5239592	1303200	11781
CL949	*S. kneri*	Shunchang, Fujian; Minjiang River	MJ 949	–	21746376	5416704	13699
CL954	*S. kneri*	Lake Qiandao, Zhejiang; Qiantangjiang River	QT 954	Aug 2005	23051680	5745336	13706
CL957	*S. kneri*	Lake Qiandao, Zhejiang; Qiantangjiang River	QT 957	Sept, 2005	10642184	8517722	13702
CL839_1	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-1	Oct 2015	6497054	5136100	11862
CL839_3	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-3	Oct 2015	8118608	6300110	12373
CL839-4	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-4	Oct 2015	6440994	4934930	11728
CL839_5	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-5	Oct 2015	7615900	5877168	12115
CL839_6	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-6	Oct 2015	6878628	5207730	10848
CL839_8	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-8	Oct 2015	7267738	5760464	11478
CL839_9	*S. kneri*	Nanjing, Jiangsu; Yangtze River	YZ 839-9	Oct 2015	7886972	5928396	11506
CL947	*S. roulei*	Shunchang, Fujian; Minjiang River	MJ 947	–	18543936	4617042	12101
CL961-2	*S. roulei*	–	– 961-2	–	16535328	4119726	12481
CL961_3	*S. roulei*	–	– 961-3	–	23520992	5857214	12071
CL946	*S. undulata*	Lake Qiandao, Zhejiang; Qiantangjiang River	QT 946	–	20562456	5122090	13478
CL946_2	*S. undulata*	Lake Qiandao, Zhejiang; Qiantangjiang River	QT 946_2	–	14934496	3718414	11567
CL944_4	*S. scherzeri*	Nanping, Fujian; Minjiang River	MJ 944_4	Aug 2005	35644536	8882916	13834
CL960	*S. scherzeri*	Lake Dongting, Hunan; Yangtze River	YZ 960	Jun 30, 2005	2385184	595500	6959
CL934_1	*S. loona*	Yangshuo, Guangxi; Pearl River	PR 934-1	Apr 2006	25821400	6432892	13950
CL934_2	*S. loona*	Yangshuo, Guangxi; Pearl River	PR 934_2	Apr 2006	10929030	8804618	12098
CL937_2	*S. obscura*	Fujian; Minjiang River	MJ 937-2	Mar 2006	8165236	6518180	11820
CL937_3	*S. obscura*	Fujian; Minjiang River	MJ 937_3	Mar 2006	8484384	6485200	11084
CL939-1_2	*S. obscura*	Shunchang, Fujian; Minjiang River	MJ 939-1_2	Mar 2006	9942692	7869080	12517
CL831_1	*C. whiteheadi*	Sanfang, Guangxi; Pearl River	PR 831_1	Sept, 2015	11168216	8776824	11862
CL935_1	*C. liui*	Shunchang, Fujian; Minjiang River	MJ 935-1	Aug 2005	21394912	5318962	13285
CL940_1	*C. whiteheadi*	Baisha, Hainan; Nandujiang River	ND 940-1	May 15, 2006	28135368	7003488	12932
CL940_2	*C. whiteheadi*	Baisha, Hainan; Nandujiang River	ND 940_2	May 15, 2006	5586960	1388952	11424
CL945_3	*C. liui*	Lake Qiandao, Zhejiang; Qiantangjiang River	QT 945_3	–	22396200	5578106	12481
CL958_1	*C. liui*	Lake Qiandao, Zhejiang; Qiantangjiang River	QT 958_1	–	30100336	7495112	14050
Out-group							
CL173	*P. nigromaculatus*	Platte River, USA	Centrarchidae	2006	16520760	4115292	11991
CL626	*M. salmoides*	Lingang, Shanghai (from aquaculture farm)	Centrarchidae	2015	18290192	4554282	12369
CL202	*P. waigiensis*	–	Latidae	–	14107776	3507554	3287
CL212	*N. spinosus*	–	Serranidae	–	18754920	4673430	11544
CL647_2	*E. awoara*	Hainan	Serranidae	2015	18139248	4518010	11357
CL301	*D. labrax*	–	Moronidae	–	12991504	3234166	11239
CL302	*P. trucha*	–	Percichthyidae	–	13482752	3356498	11982
CL322	*L. japoricus*	Luchaogang, Shanghai	Lateolabracidae	2012	16618472	4138728	12231
CL8-2	*M. swinhonis*	Lake Nanyihu, Langxi, Anhui	Odontobutidae	2009	10309292	10170586	12557
1100[d]	*E. acanthopoma*	–	Eleotridae	–	8574974	8344860	11780

"–" means the lack of information.
[a] The total number of the raw reads.
[b] The number of reads after removing PCR-duplicates and low quality reads.
[c] The number of captured loci.
[d] DNA sample from Texas A&M University - Corpus Christi.

japonicus (Lateolabracidae), *Percichthys trucha* (Percichthyidae), *Morone labrax* (Moronidae), *Niphon spinosus*, *Epinephelus awoara* (Serranidae), *Psammopperca waigiensis* (Latidae), *Micropercops swinhonis* and *Eleotris acanthopoma* (Gobioidei), which covered most putative sister taxa and closed relatives that have been proposed in previous studies. Genomic DNA was extracted from ethanol-preserved fin or muscle tissue using Ezup Column Animal Genomic DNA Purification Kits (Sangon, Shanghai, China). The concentration of the extracted DNA was quantified using a NanoDrop 3300 Fluorospectrometer (Thermo Fisher Scientific, Waltham / Wilmington, MA/DE, USA). Voucher specimens were lodged at the fish collection of Shanghai Ocean University or otherwise noted in Table 1.

2.2. Target gene markers, baits design and synthesis

Single-copy nuclear gene markers were selected using EvolMarkers tool (Li et al., 2012). Briefly, single-copy coding sequences were extracted from eight well-annotated fish genomes including: *Lepisosteus oculatus*, *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus*, *Tetraodon nigroviridis*, *Anguilla japonica*, *Gadus macrocephalus*, and *Oreochromis niloticus*, following EvolMarkers pipeline (Li et al., 2012). A suite of 17,817 single-copy conserved nuclear coding markers was found. The sequences of *O. niloticus* (3,508,171 bp) were used to design 120 bp biotinylated RNA baits (MYcroarray, Ann Arbor, Michigan). The bait sequence was padded with thymine (T) to 120 bp at the 3′end for those targets that were

less than 120 bp. The information of the selected markers, and the final bait sequences can be found in supplemental materials.

2.3. Library preparation, gene capture and sequencing

Genomic DNA was sheared to approximately 250 bp using a Covaris E220 Focused-ultrasonicator (Covaris, Woburn, USA). Subsequently, 350–500 ng of the sheared DNA from each sample was used to construct libraries. Gene capture was performed on each sample library following the manual of the MYBaits Target Enrichment System (MYcroarray, Ann Arbor, Michigan) with some modification (Li et al., 2013). To increase the number of captured genes, double capturing for each sample was performed according to the protocol (Li et al., 2013). Custom 8 bp indices on the adapter were used to distinguish reads from different samples. After gene capture, one hundred indexed samples were pooled consisting of 33 samples of this study and 67 samples from other projects in equimolar quantities for sequencing on one Illumina HiSeq 2500 lane (Illumina, Inc, San Diego, CA).

2.4. Data assembly and alignment

The raw reads were parsed according to the custom 8 bp indices for different samples. A wrapper, trim_galore v0.4.1 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, accessed on June 24, 2016) was used with Cutadapt (Martin, 2011) to trim adapter sequence and exclude reads with Phred score less than 20. The subsequent downstream assembly and alignment were performed following the methods described in Yuan et al. (2016). Firstly, PCR amplification duplicates were removed, and then the reads were parsed into different files according to their similarity to each target sequence of the query (*O. niloticus*) using a custom perl script (Yuan et al., 2016). Then, *de novo* assembly was performed to directly reconstruct reads into contigs without a reference using Trinity (Grabherr et al., 2011). The resulting output containing more than one contig were processed with Geneious V7.1.5 (Kearse et al., 2012) to merge overlapping contigs from the same target region. Smith-waterman algorithm (Smith and Waterman, 1981) was used to find the best matched sequence by comparing the query sequence with contigs derived from *de novo* assembly. Finally, putative orthologous gene was chosen by aligning the best matching contig to the genome of *O. niloticus* using blast v2.2.27 (Camacho et al., 2009). If the alignment returned a hit located out of the target region, the contig was discarded. The final output included three files: one containing coding region without flanks, one containing coding region with flanks and the other one containing intron sequence. The sequences without flanking regions were used for subsequent analyses.

2.5. Phylogenetic analysis

Each individual locus was aligned using Clustal Omega v1.1.1 (Sievers et al., 2011) with default parameter settings. To test the effect of missing data on phylogenetic reconstruction, three datasets: loci with no missing taxa (786 loci, 262,458 bp), each locus with up to 10% missing taxa (7011 loci, 1,633,572 bp), and all data included (16,943 loci, 3,213,258 bp) were analyzed in RAxML v8.0.0 (Stamatakis, 2006, 2014) and ExaML v3 (Kozlov et al., 2015) under GTRGAMMA model. The alignment of all three datasets was partitioned by codon position. A thousand bootstrap replicates were performed to assess the nodal supports. PartitionFinder v2.0 (Lanfear et al., 2012) was used to select best-fit partitioning schemes to compare with the analyses applying partitioning by codon position on the 786 loci dataset. The best partitioning scheme was picked under GTRGAMMA model and BIC criterion using the relaxed clustering algorithm to 0.01% (i.e.--rcluster-percent 0.01) (Lanfear et al., 2014).

Individual gene trees were inferred using RAxML v8.0.0 with the GTRGAMMA model. All individuals of each species were used to reconstruct the gene tree. Gene trees were summarized into a species tree with ASTRAL 4.10.6 (Mirarab et al., 2016, 2014; Mirarab and Warnow, 2015). ASTRAL is statistically consistent and has outstanding accuracy improving coalescent-based and concatenation methods with certain amounts of incomplete lineage sorting. All three datasets, 786 loci, 7011 loci and 16,943 loci were analyzed separately to test the effect of missing data on species tree reconstruction. Quartet support setting, -t 3 was chosen in the ASTRAL analyses.

2.6. Single nucleotide polymorphisms (SNPs) extraction and species delimitation

To facilitate single nucleotide polymorphisms (SNPs) calling among multiple samples, consensus reference sequences were made from coding sequences of closely related taxa of interest with no missing data (Yuan et al., 2016). BWA mapping was performed to align the trimmed fastq sequence reads back to the consensus reference implemented in BWA v0.7.5a (Li and Durbin, 2009) with default parameter settings. Alignments with the sequence map (SAM) format (Li et al., 2009) were converted into binary (BAM) format. BAM files were processed for SNPs and indel detection using GATK v 3.2.2 (DePristo et al., 2011; McKenna et al., 2010) with standard hard filtering parameters. The core NGS data processing steps and the key methods of variant discovery followed Van der Auwera et al. (2013). One of the best SNP site with fewest missing taxa and highest SNP calling score was chosen for each target region to meet the requirement of linkage equilibrium for subsequent analyses. VCF files from SNP calling results were converted to Nexus format for Bayes factor delimitation analysis (BFD*; Bouckaert et al., 2014) using a custom Perl script (Yuan et al., 2016).

To determine the number of species in *Coreoperca*, we tested two scenarios with genome-wide SNP data using the program SNAPP (Leaché et al., 2014) and BFD* (Bouckaert et al., 2014). One scenario contained two samples collected from Qiantangjiang River basin in Zhejiang, where the type specimens of *C. liui* were found, and two samples collected from the Nandujiang River in Hainan Island, where the holotype of *C. whiteheadi* was found. In total, 7508 SNPs were obtained from alignments of the four *Coreoperca* samples in addition to two outgroups (*S. roulei* and *S. obscura*). In the other scenario, we added two samples of *Coreoperca*, each collected from the Minjiang River (Fujian province) and the Pearl River (Guangxi province) respectively. Following our phylogenetic results, the specimen from the Minjiang River was lumped with samples from the Qiangtangjiang River, and the specimen from the Pearl River was grouped with samples from the Nandujiang River, which resulted in 6874 SNPs. In both scenarios, we compared the model by lumping all *Coreoperca* samples into one species or separating them into two putative species, *C. whiteheadi* and *C. liui*. We conducted path sampling with 48 steps to estimate the marginal likelihood with a Markov chain Monte Carlo (MCMC) chain length of 200,000 and a pre-burnin of 50,000 following the recommended settings in BFD* (Leaché et al., 2014). The strength of support for compared hypotheses was evaluated from Bayes factor scale, $2ln(BF)$ using the framework of Kass and Raftery (1995). The BF scale is as follows: $0 < 2ln(BF) < 2$ is not worth more than a bare mention, $2 < 2ln(BF) < 6$ means positive evidence, $6 < 2ln(BF) < 10$ represents strong support, and $2ln(BF) > 10$ represents decisive support.

To test whether *S. chuatsi* and *S. kneri* are genetically distinct and whether *S. obscura* and *S. loona* are two valid species, we

conducted similar analyses as described above. We extracted 4784 SNPs from alignments of 5 *S. chuatsi*, 10 *S. kneri* and 2 outgroup samples (*S. roulei* and *S. obscura*), and 5864 SNPs from *S. obscura*, *S. loona*, *S. undulata* and an outgroup samples (*S. roulei*). The models lumping *S. chuatsi* and *S. kneri* as one species or grouping *S. obscura* and *S. loona* as one species were compared to model assigning them as separate species.

2.7. Time calibration

Three subsets of the data, each containing 100 randomly selected loci from 6086 loci without missing any of the 11 taxa, including nine ingroups and two sister taxa were used to calibrate the time tree. One individual from each putative species was chosen to fit speciation model. Subsampling also was necessary to make analyses tractable and to obtain well-mixed runs in BEAST analyses (Bouckaert et al., 2014). BEAUti v2.3.2 was used to generate the BEAST input XML files (Drummond and Rambaut, 2007). Data were partitioned by three codon positions to accommodate rate heterogeneity. Substitution model used was GTRGAMMA with four rate classes and unlinked among partitions. A relaxed clock lognormal model and a birth-death speciation model with default priors were applied and linked across partitions. For the relaxed clock lognormal model, two parameters, mean and standard deviation were given exponential distribution. For the birth-death speciation model, speciation rate and extinction rate as uniform prior were implemented. The estimated RAxML tree was converted into a time calibrated tree based on fossil dates using penalized likelihood in R, which was used as a starting tree in the BEAST calibration. The subtree-slide, Wilson-Balding, and narrow and wide exchange operators were turned off to prevent BEAST from exploring topology space. Two replicate analyses of 20,000,000 generation of the Markov chain Monte Carlo (MCMC) with 4,000,000 preburn-in were carried for each of the 100-loci datasets. The log files were examined in Tracer v.1.6 (Rambaut et al., 2014) to verify effective sampling of all parameters and to assess convergence of independent chains. After removing 20% of samples as burn-in, a maximum clade-credibility (MCC) tree, with means and 95% highest posterior density of divergence times was constructed using TreeAnnotator v.1.8.1 (Drummond and Rambaut, 2007). The resulting trees were viewed in FigTree v1.4.0 (Rambaut, 2013).

Three calibration points were selected as priors for divergence time estimation. The hard minimum ages and soft maximum ages were assumed based on the reported fossil age of related taxa (Chen et al., 1999; Liu and Su, 1962; Liu et al., 1962). †*Tungtingichthys* was suggested as the ancestor of the sinipercids fishes based on counts of the dorsal, anal and pelvic fin rays, vertebrae, and its marginal caudal fin (Ohe, 1984). †*Tungtingichthys* was used to constrain the time to the most recent common ancestor (MRCA) of the clade of Siniperca + Coreoperca. †*Tungtingichthys was* widely found in China, and its stratigraphic horizon age estimate is Paleogene (23.03–66 Ma; Li, 1987). †*Siniperca wusiangensis* was the only fossil assigned to the genus *Siniperca*, so it was used as the minimum hard bound to constrain the MRCA of *Siniperca* (Liu and Chen, 1994). The date of †*S. wusiangensis* was estimated as from Pliocene based on paleomagnetic method, (2.39–3.3 Ma; Cao et al., 1985). †*Coreoperca shandonggensis* was a fossil of the genus *Coreoperca* and its stratum age was estimated at 10–16 Ma (Yang and Sun, 2000), so it was used to calibrate the MRCA of *Coreoperca*. The fossil otoliths assigned to the *Acanthomorphorum forcallensis* (124–122 Ma) was used to constrain the maximum age of the clade of Sinipercidae (Nolf, 2004; Santini et al., 2009). All fossil priors were placed under a log-normal distribution, allowing for the possibility that these nodes are considerably older than the fossils themselves. A hard minimum age, 23.03 Ma and 95% soft maximum age, 124 Ma (mean, 3.3; standard deviation, 0.8) were applied to the ancestral node of *Siniperca* and *Coreoperca*. A hard minimum age, 10 Ma and a 95% soft maximum age, 66 Ma (mean 2.71, standard deviation 0.8) were applied to the ancestral node of *Coreoperca*. The ancestor (crown) of *Siniperca* was constrained with a hard minimum age, 2.39 Ma and a 95% soft age, 66 Ma (mean 2.38, standard deviation 0.8).

2.8. Reconstruction of ancestral distribution areas of the sinipercids

Lagrange Analysis (DEC model; Ree and Smith, 2008) and Statistical Dispersal-Vicariance Analysis (S-DIVA; Nylander et al., 2008; Yu et al., 2010) implemented in RASP v3.1 (Yu et al., 2015, 2010) were used to reconstruct the possible ancestral ranges of the sinipercids. The trees sampled from time-calibrated analyses were used. After removing 20% of samples as burn-in, 16,000 trees from MCMC output were used to run DEC and S-DIVA. For DEC model, dispersal constrains were set so that gene flow between non-adjacent areas were not allowed. The number of maximum areas was kept as six. Outgroups were excluded from the analysis. Meanwhile, sinipercids biogeography was evaluated with six different models (DEC, DEC+j, DIVALIKE, DIVALIKE+j, BAYAREALIKE, BAYAREALIKE+j) implemented in BioGeoBEARS for comparison (Matzke, 2013). The number of maximum area and the input tree used was the same as analyses in RASP.

Six biogeographic areas were defined based on major river basins and distributions of extant sinipercids: A. the Pearl River and the Nandujiang River; B. independent rivers that flow in Fujian and Zhejiang province; C. the Yangtze River and the Huaihe River; D. the Yellow River and the East Liaohe River; E. the Yalujiang River and the Amur River; F. the Korea Peninsula (Li, 1981; Zhou et al., 1988). In our study, only *C. whiteheadi* was found in the Nandujiang River in Hainan island (Li, 1991), which was connected to the main island in recent time (Li, 1981), so we grouped the Pearl River and the Nandujiang River as zone A. Our phylogenetic analysis suggested that zone B consisting of the Qiantangjiang River and the Minjiang River would be appropriate. The Yangtze River once connected to the Huaihe River as channels to the ocean from the Palaeocene epoch to the early Oligocene (Li, 1981). The Yellow River and the Liaohe River had connected to each other; whereas the Yalujiang River and the Amur River had a river capture event (Li, 1981).

3. Results

3.1. Sequencing results and assembly

Illumina sequencing returned in an average of 14,891,408 reads per sample (2,385,184–35,644,536). After adapter trimming and quality filtering, we obtained 5,395,117 reads per sample (595,500–10,170,586). The raw reads were lodged in NCBI Sequence Read Archive (SRA) with accession numbers SRR5074315, SRR5091908, SRR5091909 and SRR5091911. *Micropercops swinhonis* had the most number of reads and *P. waigiensis* had the least number of reads. Of the 17,817 target loci, 16,943 were captured in at least in one sample. In average, 11,965 loci were captured per sample. Only *P. waigiensis* and one *S. scherzeri* (CL960) had fewer loci captured probably due to their low DNA quality (Table 1). The average length of locus was 190 bp and totaled 3,213,258 bp for the 16,943 loci.

3.2. Phylogenetic relationships

Our concatenated ML analyses resulted in a well-resolved phylogeny that is consistent when datasets containing 786, 7011, or 16,943 loci were used (Fig. 2; Figs. S4–S6). All species were

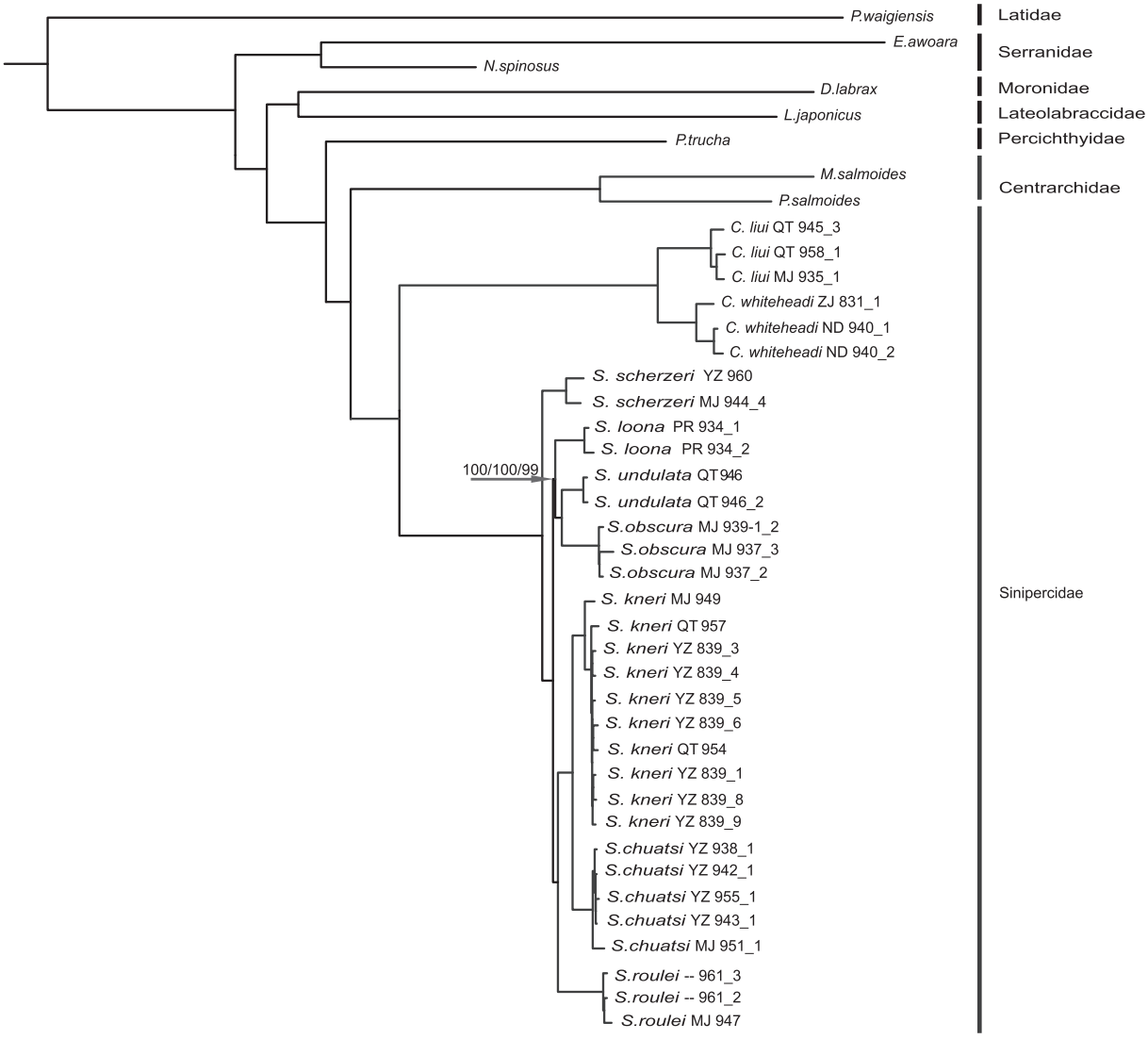

Fig. 2. The concatenated ML tree of the sinipercids based on three datasets: loci have no missing taxa (786 loci, 262,458 bp), each locus could have up to 10% missing taxa (7011 loci, 1,633,572 bp), and all data included (16,943 loci, 3,213,258 bp). All clades are supported with 100 bootstrap values except that the clade of *S. loona*, *S. undulata* and *S. obscura* has a support value of 99 for the 786 loci dataset. Gobioidei was used as outgroup (not shown). For sample code after the species name, see Table 1.

monophyletic in the resulting phylogeny with bootstrap support value of 100 (Fig. 2). All internal nodes also had high bootstrap support values of 100 except that in the analysis using 786 loci, where the node unifying *S. loona*, *S. obscura* and *S. undulata* had a bootstrap value of 99. The best partitioning scheme suggested by PartitonFinder had 45 partitions for the 786 loci data (supplementary file: PartitionFinder scheme.txt), with a resulting phylogeny similar to the result of analyses partitioned by codon position (Fig. S7).

Species tree reconstructed using ASTRAL was consistent with the results from the ML analysis with a normalized quartet score 0.7. Analyzing the three different datasets also produced exactly the same result. All nodes had a bootstrap value of 100, except that the species tree based on the 786 loci data had three nodes with bootstrap values ranged between 96 and 99 (Fig. 3). We found that the Centrarchidae (represented by *M. salmoides* and *P. nigromaculatus*) is sister to the Sinipercidae, and the Percichthyidae (*P. trucha*) is sister to a clade consisting of the Centrarchidae and the Sinipercidae. The Sinipercidae consists of two monophyletic genera, *Siniperca* and *Coreoperca*. Within *Siniperca*, *S. scherzeri* is most basal. The rest species were split into two groups: one clade grouping *S. chuatsi*, *S. kneri* and *S. roulei* together and the other clade containing *S. obscura*, *S. loona* and *S. undulata* (Figs. 2 and 3).

3.3. Species delimitation

Bayesian species delimitation analysis provided strong supports for all three species pairs to be genetically distinctive. The Bayes factor scores were decisive in splitting samples of *Coreoperca* collected from the Nandujiang River and the Qiantangjiang River (2*ln*BF = 2646, Table 2). When the samples from the Pearl River and the Mingjiang River were added, two species of *Coreoperca* were still highly supported (2*ln*BF = 14,648, Table 2). Similarly, we found that *S. chuatsi* and *S. kneri* are two genetically different species instead of one (2*ln*BF = 62,014, Table 2) and *S. obscura* and *S. loona* also are distinct (2*ln*BF = 39,728, Table 2). The splitting of *S. loona* and *S. undulata* is also supported (2*ln*BF = 18,708, Table 2).

3.4. Divergence time analyses

Two independent runs for each of the three 100-loci dataset reached stationarity in BEAST v2.3.2. Effective sample sizes for all parameters were above 200 and all analyses produced similar results. The age of the common ancestor of all sinipercids was estimated at 53.1 Ma (CI: 30.4–85.8 Ma; Fig. 4) during the early

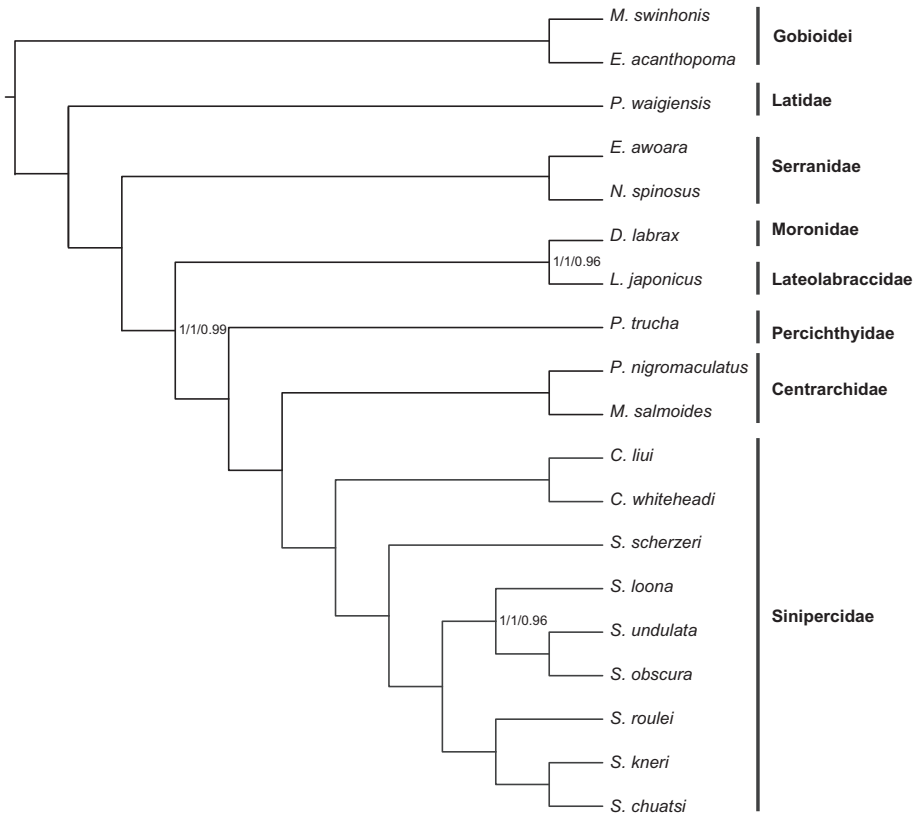

Fig. 3. The species tree obtained from ASTRAL analysis on the three datasets as described in Fig. 2. The number on braches are bootstrap supports for analyses based on 16943/7011/786 loci.

Table 2
Results of species delimitation based on SNPs data using BFD*.

Group	Model	Marginal likelihood	2lnBF
C. whiteheadi and *C. liui*	Splitting samples from QTJ and NDJ	−9662	
	Lumping individuals from QTJ and NDJ	−10985	2646
	Splitting two clades of *Coreoperca*	−15403	
	Lumping two clades of *Coreoperca*	−22727	14648
S. chuatsi and *S. kneri*	Splitting *S. chuatsi* and *S. kneri*	−28755	
	Lumping *S. chuatsi* and *S. kneri*	−59762	62014
S. loona, *S. obscura* and *S. undulata*	Splitting *S. loona*, *S. obscura* and *S. undulata*	−22466	
	Lumping *S. loona* and *S. obscura*	−42330	39728
	Lumping *S. obscura* and *S. undulata*	−37467	30002
	Lumping *S. loona* and *S. undulata*	−31820	18708

Eocene. The ancestor of genus *Siniperca* was estimated at 12.2 Ma (CI: 6.5–19.8 Ma) during Miocene. The node splitting the *S. chuatsi* clade vs. *S. obscura* clade was estimated to be 9.4 Ma (CI: 5.2–15.5 Ma). Within the *S. chuatsi* clade, *S. roulei* split off at 8.4 Ma (CI: 4.5–14.2 Ma) and *S. chuatsi* and *S. kneri* diverged from each other at 4.3 Ma (CI: 1.7–7.5 Ma). In the clade of *S. obscura*, *S. loona* first split off at 8.7 Ma (CI: 4.5–14.4 Ma) and then *S. obscura* and *S. undulata* diverged at 7 Ma (CI: 3.4 – 12 Ma). The two *Coreoperca* lineages were separated at 15 Ma (CI: 10.9–22.3 Ma, Fig. 4).

3.5. Biogeographical analysis

The results of DEC analysis showed that the ancestor of all sinipercids was widely distributed in southern China, including areas south of the Yangtze River (A, B, C or AB, AC, BC in Fig. 5). The results of S-DIVA analysis showed even wider range of distribution of the ancestor of all sinipercids (Fig. S8). Two ancestor distribution areas were recovered with similar possibilities for *Coreoperca* living in China (AB, AC in Fig. 5). Because of the lack of the samples of *Coreoperca* from Japan and Korea, the ancestral area of all *Coreoperca* cannot be established. The genus *Siniperca* has an ambiguous ancestral area reconstruction and four origins are found (Fig. 5), thus, it might distribute widely in southern China. The ancestor of *S. roulei*, *S. chuatsi*, and *S. kneri* and the ancestor of *S. chuatsi* and *S. kneri* have equal probability distributing in one of the zone in southern China (A, B, C in Fig. 5). In BioGeoBEARS analysis, the best model suggested by AIC value is BayArea without *j* parameters. The results of BioGeoBEARS analysis are similar as the DEC results implemented in RASP (Fig. S9).

4. Discussion

4.1. Species delimitation

The genus *Coreoperca* used to have three described species: *C. whiteheadi*, *C. herzi* and *C. kawamebari* (Zhou et al., 1988). *Coreoperca kawamebari* is distributed mainly in Japan and part of Korea, whereas *C. herzi* is distributed only in Korea (Kim et al., 1997; Masuda et al., 1984). *Coreoperca whiteheadi* is endemic to China and northern Vietnam (Gao, 1991; Kottelat, 2001), and it was once the only described species in China until Cao and Liang (2013) described a new species, *C. liui* based on 19 samples collected from

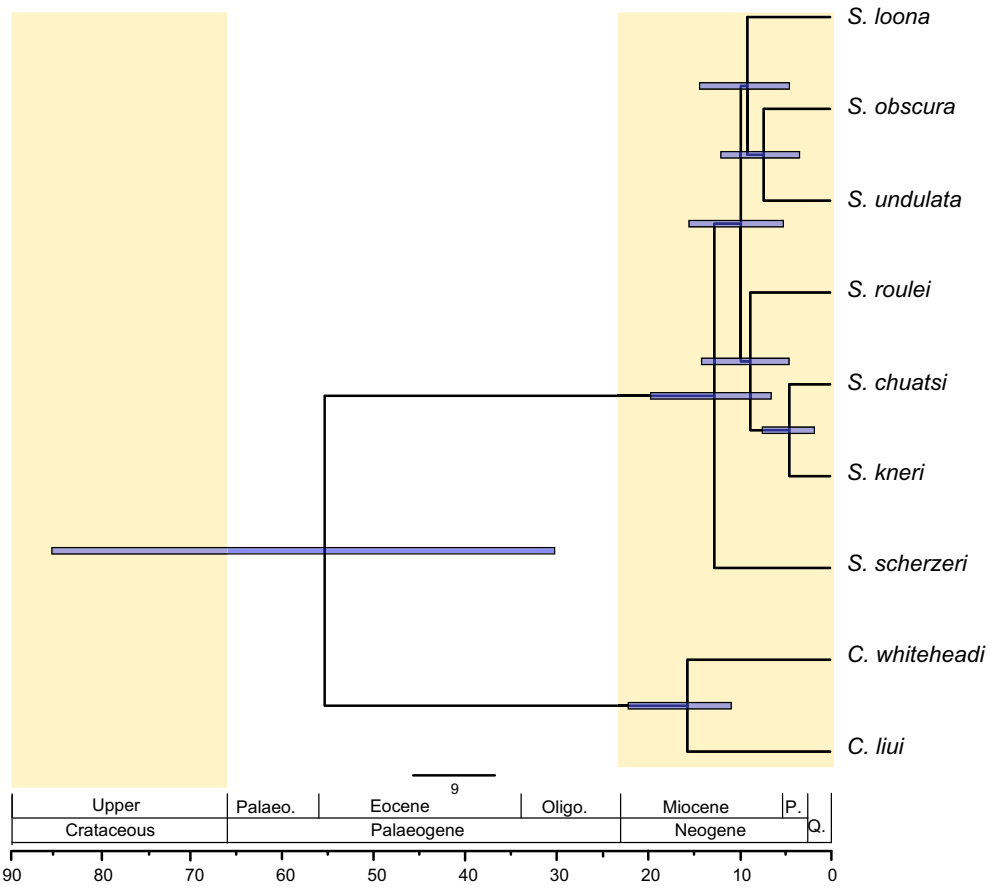

Fig. 4. Time-calibrated phylogeny of the sinipercids inferred from BEAST based on 100 loci. Mean divergence time estimates are shown with 95% highest posterior density (HPD; blue bars). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

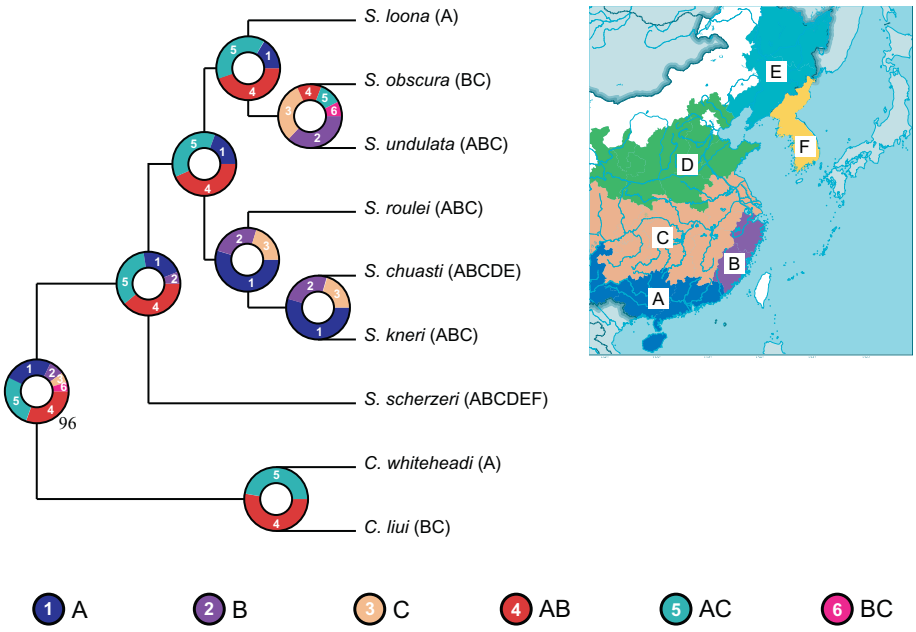

Fig. 5. Result of Lagrange Analysis (DEC model) on the sinipercids. Biogeographic zone: A. The Pearl River and the Nandujiang River; B. Independent rivers flow Fujian and Zhejiang province; C. The Yangtze River and the Huaihe River; D. The Yellow River and the East Liaohe River; E. The Yalujiang River and the Amur River; F. The Korea Peninsula. Pie charts at the nodes indicate probabilities supporting for particular ancestral range reconstructions. Outgroups *Micropercops swinhonis* and *Eleotris acanthopoma* are excluded from the analysis.

the Qiantangjiang River. The characters distinguishing this new species from *C. whiteheadi* included different body depth (depth 28.3–31.6% vs. 33.5–39.1% SL) and eye size (diameter 4.6–5.6% vs. 6.6–8.5% SL) than those of *C. whiteheadi*. However, the validity of the new species had not been tested explicitly using genetic data. Our results showed that the samples collected from the Nandujiang River and the Qiantangjiang River diverged from each other 15 million years ago, older than ages of all the other *Siniperca* species (Fig. 4). BFD* analysis also supported the two Chinese *Coreoperca* lineages as different species (2LnBF = 2646, Table 2). Our sample from the Minjiang River is nested within samples collected from the Qiantangjiang River and the sample collected from the Pearl River is grouped with samples from the Nandujiang River (Fig. 2). Cao et al. (2013) found two clades of *Coreoperca* separated by the Nanling-Wuyi Mountain range based on mtDNA cytb and S7 intron1 data, but they did not sample the Nandujiang River, where the holotype of *C. whiteheadi* was described. Zhao and Zhao (2007) also found two clades of *Coreoperca* based on mitochondrial control region sequences, their results showed that samples from Pearl River were different from samples collected from the Nandujiang River, but they were very similar to the samples from the Yangtze River, which is inconsistent with our results, probably due mitochondrial-nuclear discordance. Nonetheless, all studies found low within population genetic diversity and isolation between different populations, suggesting that more cryptic species might exist in *Coreoperca*. *Coreoperca* lives in small streams of mountains with disjunctive distribution area (Zhou et al., 1988), thus it might restrict gene flow between different populations and drove speciation, therefore, samples from wider distribution areas should be targeted for future study to further elucidate species status in *Coreoperca*.

Siniperca chuatsi was distributed from the Pearl River and northward up to the Amur River, but *S. kneri* was absent to the north of the Huaihe River. *Siniperca chuatsi* was distinguished from *S. kneri* in that its maxilla extends beyond the rear end of eye, whereas the maxilla of *S. kneri* never passes over the rear end of its eyes (Zhou et al., 1988). They were two distinct species in morphology and distribution pattern, but mitochondrial data could not separate them because *S. kneri* was found embedded in the clade of *S. chuatsi* (Zhao et al., 2008). Our BFD* analyses with a decisive support suggested that *S. chuatsi* and *S. kneri* are two distinct species (Table 2). The reason for the mitochondrial data being indecisive could be due to incomplete sorting between the two recently splitting species or due to introgression. With thousands of independent loci, our analyses for the first time provided firm evidence for that *S. chuatsi* and *S. kneri* are two distinct species.

Siniperca loona was thought to be synonymous to *S. obscura* (Zhou et al., 1988), but Liu and Chen (1994) argued that *S. loona* should be considered as a valid species because it has scaly cheeks and lower body height than that of *S. obscura* and it was distributed in the Pearl River, middle and upper reach of the Yangtze River, while *S. obscura* was distributed in lower reach of the Yangtze River and southeastern part of China. Kong and Zhou (1993) found minor differences between *S. obscura* and *S. loona* based on comparison of skeletal characteristics. Zhao et al. (2007) found *S. loon* and *S. obscura* as distinct clades based on mitochondrial control region data. Our phylogenetic results and BFD* species delimitation supported *S. loona* to be a valid species (Fig. 2, Table 2).

4.2. Phylogenetic interrelationship of the sinipercids

Most of the previous morphological and molecular studies agreed with that the Sinipercidae is monophyletic (Li et al., 2010; Liu, 1997; Liu and Chen, 1994; Near et al., 2012; Zhao et al., 2008, 2006a, 2006b). The concatenated ML analyses and coalescent species trees analysis resulted in an exactly, strongly supported tree with all clades had BS values of 100%. Tonini et al. (2015) suggested that concatenation method could perform as well or better than species tree method. Our study concurred with the suggestion that both concatenation method and species tree method are relevant and should be compared in phylogenetic study (Tonini et al., 2015). Our results supported the sinipercids as a monophyletic group (Figs. 2 and 3). Previous classifications placed the sinipercids within the families Serranidae (Jordan, 1923; Nichols, 1943; Zhou et al., 1988), Percichthyidae (Gosline, 1966; Liu and Chen, 1994; Nelson, 1976, 1984), Centropomidae (Nelson, 1994; Waldman, 1986), and Sinipercidae (Nelson, 2006; Roberts, 1993) or as sister taxon to the centropomids, centrarchids, and *Lateolabrax* (Liu, 1997) or close to *Perca* and *Pristigenys* (Zhao et al., 2005). Our phylogenetic analyses placed the centrarchids as the sister group of the sinipercids with BS values of 100%.

Zhou et al. (1988) divided the sinipercids into three genera: *Coreoperca*, *Siniperca* and *Coreosiniperca* based on morphological characters of a wide collection of fishes from China, Japan and Korea, which was corroborated by Kong and Zhou (1993) after examining skeletal characteristics of seven species of the sinipercids. However, Liu and Chen (1994) suggested that the Sinipercidae consisted only of two genera: *Siniperca* and *Coreoperca* on the basis of their cladistic analysis on 34 osteological characters, and assigned *Coreosiniperca roulei* to *Siniperca*, which was also supported by subsequent molecular studies (Fig. 1; Chen et al., 2010; Li et al., 2010; Zhao et al., 2006a, 2006b). We recovered two major clades: *Siniperca* and *Coreoperca*, which is consistent with the previous studies (Chen et al., 2010; Li et al., 2010; Zhao et al., 2006a, 2006b). Nevertheless, we identified some new relationships. For example, we found *S. scherzeri* as the oldest *Siniperca* species, which was only supported by Li et al. (2010), while many other authors found *S. scherzeri* was more closely related to derived sinipercids (Fig. 1). A basal *S. scherzeri* is more consistent with its current distribution range than a derived *S. scherzeri*, because its distribution extends into the Korean Peninsular, which is compatible with an old *S. scherzeri* lineage (more discussion see Section 4.3). Another unexpected result is that *S. roulei* was found closely related to *S. chuatsi* and *S. kneri* with strongly supported in all analyses, which has not been reported before.

4.3. Divergence time and historical biogeography of the sinipercids

The Sinipercidae was widely dispersed in the East Asia, extending from Amur River in the north to Nandujiang River, Hainan Island, China in the south. Eastward, it expanded to the southern part of the west side of Japanese Honshu island, and westward to the lower reaches of the Jinshajiang River of the west Sichuan basin in China (Li, 1991). *Siniperca chuatsi* and *S. scherzeri*, which span the Palaearctic and the Oriental, were the most widely distributed sinipercids. Other species of *Siniperca* and the Chinese *Coreoperca* were distributed in many river systems of South China, preferring warm temperate but their reproduction needed some degrees of low-temperature in the winter (Li, 1991). Most sinipercids occur below 500 meters above sea level excepted that *S. scherzeri* has been found in Lake Qionghai at 1767 meters in the west Sichuan basin (Li, 1991; Zhou et al., 1988), which was attributed to presumably very recent rising up of the region and suitable conditions in the lake to *Siniperca* (Li, 1991). *Siniperca chuatsi*, *S. kneri* and *S. scherzeri* inhabit large rivers and lakes, whereas *S. obscura*, *S. loona*, *S. undulata*, *S. roulei* and *Coreoperca* were often found in tributaries or in relatively small drainage (Chen et al., 1999).

Our divergence time analysis provided a robust time-scale for the Sinipercidae originating around 53.1 Ma (Fig. 4). It is the first time that fossil evidences of the sinipercids were used to calibrate the time tree of the Sinipercidae, and the inferred time of their origin are older than estimates from previous studies (Chen et al.,

2004; Zhao et al., 2008). Warm and wet climatic conditions prevailed over northern latitudes, even within the Arctic, for much of the Tertiary and especially during the Eocene (54–34 Ma) with cooling gently and then fluctuated until 15 Ma with cooling progressively (Guo, 2010; Milne and Abbott, 2002). The climate condition of the geologic periods was most favorable to the sinipercids at the origin while the later cooling and fluctuation might drive speciation events in the sinipercids.

The age of the split leading to the separation of *Siniperca scherzeri* from the genus *Siniperca* was estimated at 12.2 Ma (Fig. 4). Speciation in other species of the genus *Siniperca* mainly occurred between late Miocene to the early Pliocene (Fig. 4). The initial collision between India and Asia in the early Palaeogene (50–55 Ma) or even earlier, about 70 Ma (Yin and Harrison, 2000) affected the major tectonic episode continued through the Oligocene and well into the Miocene (Harrison et al., 1992). During these times, associated geological processes occurred, ranging from the uplift (thickening) of the Himalaya–Tibetan plateau to lateral extrusion of the continental landmass. The Himalaya orogeny caused the tectonic evolution of China, including faulting and subsidence of east and uplifting of west of Wulingshan Mountain in south China (Jia et al., 2004). Geological researches indicated that the southern and central Tibetan plateau probably has reached the highest elevation in the Middle Miocene (about 14 Ma; Blisniuk et al., 2001; Coleman and Hodges, 1994), or even earlier (at 15 Ma; Spicer et al., 2003), suggesting that southern China had begun to be established geomorphologic feature at least as early as Middle Miocene time. The Asian Monsoon formation occurred before 22 Ma (Guo, 2010) and then continued intensification, about 8.5–6 Ma ago (Guo, 2010; Kroon et al., 1991), influencing rainfall and rain belt location. Geological change and rainfall variability might cause habitat diversification and promote sympatric differentiation of the species in genus *Siniperca*.

Siniperca scherzeri is widespread in south China, but occurs in scattered localities in Korea and Liaoning province. The Changbai Mountain Range is located in eastern of the Northeast China, stretching along the border between China and Korea. The orogeny of the Changbai Mountain originated from volcanic eruptions in the Oligocene Epoch around 28.4 Ma, mainly occurring in the Miocene, Pliocene and the early Pleistocene, and stopped in the Pleistocene at 1.2 Ma (Wan, 2012; Yin, 2010). The current distribution of *S. scherzeri* in Korea corroborates the hypothesis of a basal *S. scherzeri* lineage older than the formation of Changbai Mountain as a barrier to its dispersal to Korea or *vice versa*, which could be one of its refugial regions for surviving cooling climate. On the contrary, the younger species, *S. chuatsi* did not expand over to the Korea peninsular. Zhao et al. (2008) pointed out *S. chuatsi* could be divided into the north (the Amur River and Yellow River) and the south (the Yangtze, Qiantang and Minjiang River) groups along the Qinling-Dabie Mountain Range and genetic diversity of the north group is significantly less than that of the south group, suggesting that the north region was colonized by the south group.

Chinese *Coreoperca* is mainly distributed from the Yangtze River, southward to Hainan Island and northeastern Vietnam in small mountainous tributaries with gravel bases (Kottelat, 2001; Li, 1991; Zhou et al., 1988). Scattered habitat limits its dispersal and gene flow, promoting population differentiation (Cao et al., 2013; Zhao and Zhao, 2007). Different river drainages and mountain range noticeably affects the population structure, and distribution of fish community. Chinese *Coreoperca* probably originated from the South China as suggested by DEC ancestor area reconstructions analysis (Fig. 5), and was split into two species at the Mean Miocene (11–22.2 Ma) coinciding with the process of the orogenic movement. These massive formations disrupted wind and weather patterns, altering rainfall distribution and Chinese topography and climate. The Himalayan orogeny and the uplift of Nanling and Wuyi in the South China created massive mountains and deeply carved valleys (Jia et al., 2004), which might act as barriers to dispersal and facilitated divergence of *Coreoperca* species. Because of the lack of samples of *C. kawamebari* and *C. herzi*, we cannot estimate the divergence time and the ancestral area of the genus *Coreoperca* as a whole, which should be done for future studies. The close relationship between *C. whiteheadi* distributed in Hainan Island and Guangxi province was observed in our phylogenetic analysis. The time of separation of the freshwater fish fauna of the Hainan Island from the continent is not well resolved (Shan, 2015). The Qiongzhou Strait, which separates Hainan Island from the continent is around 40 meters deep and could not act as a significant barrier to gene flow when accounting the descending of sea level about 100 m in Quaternary ice age (Li, 1981). More samples of *C. whiteheadi* from the continent just across the Qiongzhou Strait and probably from North Vietnam should be collect in the future to test the hypotheses about the origin and the age of the *C. whiteheadi* population of the Hainan Island.

5. Conclusions

We reconstructed a robust phylogeny of the sinipercids using thousands of nuclear coding sequences obtained through target enrichment. We confirmed the monophyly of the sinipercids and the sister-taxa relationship between the centrarchids and the sinipercids as proposed in previous studies. The results of Bayes factor delimitation (BFD*) analysis using thousands of SNPs supported *S. loona* and *C. liui* as the valid species and provided first genetic evidence distinguishing *S. chuatsi* and *S. kneri* as separate species. We also found *S. scherzeri* as the most basal *Siniperca* species, which is more consistent with current distribution of the sinipercids than a derived *S. scherzeri* as reported from previous studies. According to the results of fossil calibration analysis and reconstruction of ancestral distribution areas, the ancestor of sinipercids fishes was widely distributed in southern China during the Late Palaeogene and the early Eocene.

Acknowledgements

This work was supported by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and the Innovation Program of Shanghai Municipal Education Commission to CL, and the China Agriculture Research System (CARS-49) to JZ. We thank Mr. Jiahu Lan for helping with sample collection. The Shanghai Oceanus Supercomputing Center (SOSC) provided computational resource for data analyses.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.03.014.

References

Blisniuk, P.M., Hacker, B.R., Glodny, J., Ratschbacher, L., Bi, S., Wu, Z., McWilliams, M. O., Calvert, A., 2001. Normal faulting in central Tibet since at least 13.5 Myr ago. Nature 412, 628–632.

Bouckaert, R., Heled, J., Kuhnert, 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 Comput. Biol. 10, e1003537.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Cao, L., Liang, X., 2013. A new freshwater perch species of the genus *Coreoperca* herzenstein (Perciformes, Serranidae, Sinipercinae) from Zhejiang Province, China. Acta Zootaxonmica Sinica 38, 891–894.

Cao, L., Liang, X.-F., Tang, W., Zhao, J., 2013. Phylogeography of *Coreoperca* whiteheadi (Perciformes: *Coreoperca*) in China based on mitochondrial and nuclear gene sequences. Biochem. Syst. Ecol. 50, 223e–231.

Cao, Z.Y., Xing, L.S., Yu, Q.h., 1985. The age and boundary of magnetic strata of the Yushe formation. Bull. Inst. Geomech. CAGS 6, 143–154.
Chang, C.-H.M., 1988. Systematics of the Centrarchidae (Perciformes: Percoidei) with Notes on the Haemal–Anal–Axial Character Complex., Department of Biology. The City University of New York, New York.
Chen, D., Guo, X., Nie, P., 2007. Non-monophyly of fish in the Sinipercidae (Perciformes) as inferred from cytochrome b gene. Hydrobiologia 583, 77–89.
Chen, D., Guo, X., Nie, P., 2010. Phylogenetic studies of sinipercid fish (Perciformes: Sinipercidae) based on multiple genes, with first application of an immunerelated gene, the virus-induced protein (viperin) gene. Mol. Phylogenet. Evol. 55, 1167–1176.
Chen, L.Y., Zhao, S.Y., Mao, K.S., Les, D.H., Wang, Q.F., Moody, M.L., 2014. Historical biogeography of Haloragaceae: an out-of-Australia hypothesis with multiple intercontinental dispersals. Mol. Phylogenet. Evol. 78, 87–95.
Chen, P.-F., Liu, H.-Z., Yan, J.-X., 1999. Discovery of fossil Coreoperca (Perciformes) in China. Vert. PalAsiat. 37, 212–227.
Chen, W.J., Ortí, G., Meyer, A., 2004. Novel evolutionary relationship among four fish model systems. Trends Genet. 20, 424–431.
Coleman, M., Hodges, K., 1994. Evidence for Tibetan plateau uplift before 14 Myr ago from a new minimum age for east–west extension. Nature 374, 49–52.
Corl, A., Ellegren, H., 2013. Sampling strategies for species trees: the effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. Mol. Phylogenet. Evol. 67, 358–366.
DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Gao, G., 1991. Serranidae. In: Pan, J.-H., Zhong, L., Zheng, C.-Y., Wu, H.-L., Liu, J.-H. (Eds.), The Freshwater Fishes of Guangdong Province. Guangdong Science and Technology Press, Guangzhou, pp. 363–371. 589 pp.
Gosline, W.A., 1966. The limits of the fish family Serranidae, with notes on other lower percoids. Proc. Calif. Acad. Sci. 33, 91–112.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Guo, Z., 2010. 22-8 Ma Eolian Deposits Recorded Monsoon Evolution. Intergrative Study on the Evolution of Western China Environment. China Meteorological Press, Beijing.
Harrison, T.M., Copeland, P., Kidd, W., Yin, A., 1992. Raising Tibet. Science 255, 1663–1670.
Jia, Z., He, T., Lu, H., 2004. Episodes and geodynamic setting of Himalayan movement in China. Oil Gas Geol. 25, 121–125.
Johnson, G.D., 1984. Percoidei: development and relationships. In: Moser, H.G., Richards, W.J., Cohen, D.M., Fahay, M.P., Kendell Jr., A.W., Richardson, S.L. (Eds.), Ontogeny and systematics of fishes. American Society of Ichthyologists and Herpetologists, Special Publication. No. 1, pp. 464–498.
Jordan, D.S., 1923. A classification of fishes including families and genera as far as known. Stanford Univ. Publ., Univ. Ser. Biol. Sci. 3, 192.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kim, Y., Cho, G., Park, S., Joo, G., 1997. Fish fauna of headwater streams in Southern Region of Korea. Acta Hydrobiol. Sin. 21, 183–194.
Kimball, R.T., Braun, E.L., 2014. Does more sequence data improve estimates of galliform phylogeny? Analyses of a rapid radiation using a complete data matrix. PeerJ 2, e361.
Kong, X.-Y., Zhou, C., 1993. Comparative studies on the skeletal characteristics of seven sinipercinae fishes of China. J. Ocean Univ. Qingdao 23, 116–124.
Kottelat, M., 2001. Freshwater fishes of northern Vietnam. A Preliminary Check-list of the Fishes known or Expected to Occur in Northern Vietnam with Comments on Systematics and Nomenclature. In: Environment and Social Development Unit, E.A.a.P.R.T.W.B. (Ed.), p. 123.
Kozlov, A.M., Aberer, A.J., Stamatakis, A., 2015. ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics 31, 2577–2579.
Kroon, D., Steens, T., SR, T., 1991. Onset of monsoonal related upwelling in the Western Arabian Sea as revealed by planktonic forminifers. Proc. ODP Sci. Results 117, 257–263.
Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82.
Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–542.
Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G.J., 2013. Capturing protein-coding genes across highly divergent species. Biotechniques 54, 321–326.
Li, C., Orti, G., Zhao, J., 2010. The phylogenetic placement of sinipercid fishes ("Perciformes") revealed by 11 nuclear loci. Mol. Phylogenet. Evol. 56, 1096–1104.
Li, C., Riethoven, J.J., Naylor, G.J., 2012. EvolMarkers: a database for mining exon and intron markers for evolution, ecology and conservation studies. Mol. Ecol. Resour. 12, 967–971.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Li, S., 1981. The Division of Freshwater Fish Fauna in China. Sciences Press, Beijing.
Li, S., 1991. Geographical distribution of the Sinipercine fishes. Chin. J. Zool. 26, 40–44.
Li, Z., 1987. Discussion early Tertiary sedimentary environment in eastern China. Geol. Chem. Miner. 1, 13–25.
Liu, H., 1997. Study on systematic position of sinipercine fishes with discussion on relationships of some lower perciforms. Trans. Chin. Ichthyol. Soc. 6, 1–7.
Liu, H., Chen, Y., 1994. Phylogeny of the sinipercine fishes with some taxonomic notes. Zool. Res. 15, 1–12.
Liu, H.T., Su, T., 1962. Pliocene fishes from Yushe Basin, Shanxi. Vertebrata PalAsiatica 6, 1–25.
Liu, T.-S., Liu, H.T., Tang, X., 1962. A new percoid fish from south China. Vertebrata Palastatica 6, 121–128.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal, vol. 17, pp. 10–12.
Masuda, H., Amaoka, K., Araga, C., Uyeno, T., Yoshino, T., 1984. The Fishes of the Japanese Archipelago, vol. 1. Tokai University Press, Tokyo, Japan.
Matzke, N.J., 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front. Biogeogr. 5, 242–248.
McCully, H.H., 1962. The relationship of the Percidae and the Centrarchidae to the Serranidae as shown by the anatomy of their scales. Am. Zool. 2, 430.
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 Res. 20, 1297–1303.
Milne, R.I., Abbott, R.J., 2002. The origin and evolution of Tertiary relict floras. Adv. Botan. Res. 38, 281–314.
Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–548.
Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–52.
Near, T.J., Sandel, M., Kuhn, K.L., Unmack, P.J., Wainwright, P.C., Leo Smith, W., 2012. Nuclear gene-inferred phylogenies resolve the relationships of the enigmatic Pygmy Sunfishes, Elassoma (Teleostei: Percomorpha). Mol. Phylogenet. Evol. 63, 388–395.
Nelson, J.S., 1976. Fishes of the World. Wiley, New York.
Nelson, J.S., 1984. Fishes of the World. Wiley, New York.
Nelson, J.S., 1994. Fishes of the World. J. Wiley, New York.
Nelson, J.S., 2006. Fishes of the World. John Wiley and Sons Inc, New York.
Nichols, J.T., 1943. The Fresh-Water Fishes of China. American Museum of Natural History, New York.
Nolf, D., 2004. Otolithes de poissons aptiens du Maestrazgo (province de Castellon, Espagne orientale). Bull. Inst. Roy. Sci. Nat. Belg, Sci. Terre. 74, 101–120.
Nylander, J.A., Olsson, U., Alstrom, P., Sanmartin, I., 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: Turdus). Syst. Biol. 57, 257–268.
Ohe, F., 1984. Fossils of the genus Coreoperca from Miocene deposits in Japan. Tansuigyo (Freshwater fishes) 10, 25–32.
Rambaut, A., 2013. Figtree v 1.4.0. Available from <http://tree.bio.ed.ac.uk>.
Rambaut, A., Suchard, M., Xie, D., Drummond, A., 2014. Tracer v1.6. Available from <http://beast.bio.ed.ac.uk/Tracer>.
Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.
Roberts, C.D., 1993. Comparative morphology of spined scales and their phylogenetic significance in the Teleostei. Bull. Mar. Sci. 52.
Santini, F., Harmon, L.J., Carnevale, G., Alfaro, M.E., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Biol. 9, 194.
Shan, J., 2015. Divergent views on the formation of Hainan Island Journal of Hainan, Hainan.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.
Smith, T.F., Waterman, M.S., 1981. Identification of common molecular subsequences. J. Mol. Biol. 147, 195–197.
Smith, W.L., Craig, M.T., 2007. Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percid fishes. Copeia, 35–55.

Spicer, R.A., Harris, N.B., Widdowson, M., Herman, A.B., Guo, S., Valdes, P.J., Wolfe, J. A., Kelley, S.P., 2003. Constant elevation of southern Tibet over the past 15 million years. Nature 421, 622–624.

Stamatakis, A., 2006. Phylogenetic Models of Rate Heterogeneity: A High Performance Computing Perspective., Proc. of IPDPS2006, Rhodos, Greece.

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

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Orti, G., 2015. Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. PLoS Curr 7.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinform. 43, 11.10.1–11.10.33.

Waldman, J.R., 1986. Systematics of Morone (Pisces: Moronidae), with notes on the lower percoids., Department of Biology. The City University of New York, New York.

Wan, F., 2012. Geological and geomorphological evolution history of Changbai Mountain of Jilin Province. Jilin Geol. 31 (21-22), 91.

Yang, S.P., Sun, B., 2000. Palaeoecology of Miocene Shanwang biota in Shandong province, East China. J. Palaeogeogr. 2, 1–11.

Yin, A., 2010. Cenozoic tectonic evolution of Asia: a preliminary synthesis. Tectonophysics 488, 293–325.

Yin, A., Harrison, T.M., 2000. Geologic evolution of the Himalayan-Tibetan orogen. Annu. Rev. Earth Planet. Sci. 28, 211–280.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87, 46–49.

Yu, Y., Harris, A.J., He, X., 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. Mol. Phylogenet. Evol. 56, 848–850.

Yuan, H., Jiang, J., Jimenez, F.A., Hoberg, E.P., Cook, J.A., Galbreath, K.E., Li, C., 2016. Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. Mol. Ecol. Resour. 16 (5), 1095–1106.

Zhao, J., Li, C., Zhao, L., Wang, W., Cao, Y., 2008. Mitochondrial diversity and phylogeography of the Chinese perch, Siniperca chuatsi (Perciformes: Sinipercidae). Mol. Phylogenet. Evol. 49, 399–404.

Zhao, J., Li, S., Cai, W., Wang, W., 2005. The preliminary phylogenetic relationships of sinipercine fishes and some lower percoids inferred from 16S ribosomal DNA sequences. J. Shanghai Fish. Univ. 14, 364–369.

Zhao, J., Li, S., Cai, W., Wang, W., 2006a. Phylogenetic relationship of sinipercine fishes in East Asia based on cytochrome b sequences analysis. Acta Zool. Sin. 52, 676–680.

Zhao, J., Wang, W., Li, S., Cai, W., 2006b. Structure of the mitochondrial DNA control region of the sinipercine fishes and their phylogenetic relationship. Acta Genet. Sin. 33, 793–799.

Zhao, L., Zhao, J., 2007. Genetic variation of the mitochondrial DNA control region among 4 populations of *Coreoperca whiteheadi*. J. Shanghai Fisher. Univ. 16, 409–413.

Zhao, L., Zhao, J., Cao, Y., 2007. Genetic Differentiation of the Mitochondrial DNA Control Region of *Siniperca obscura* from Minjiang River and Lijiang River. Chin. J. Zool. 42, 54–58.

Zhou, C., Yang, Q., Cai, D., 1988. On the classification and distribution of the sinipercinae fishes (family Serranidae). Zool. Res. 9, 113–125.

Zhu, Y., 1985. The Fishes of Fujian Province (Part II). Fujian Science and Technology Press, Fuzhou, China.