Check for updates

Molecular Ecology

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE

Doubled Genomes, Divergent Fates: Genomic Insights Into Diversification in an Allotetraploid Cavefish

Santiago Montero-Mendieta[1] | Yuwei Wang[1,2] | Chongnv Wang[1] | Fanwei Meng[1] | Yahui Zhao[1] | Xinxin Li[1] | Baocheng Guo[1,2]

[1]State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing, China | [2]University of Chinese Academy of Sciences, Beijing, China

Correspondence: Xinxin Li (lixinxin@ioz.ac.cn) | Baocheng Guo (guobaocheng@ioz.ac.cn)

Received: 30 January 2025 | **Revised:** 5 September 2025 | **Accepted:** 16 September 2025

Handling Editor: Kaichi Huang

Funding: This study was supported by grants from the Ministry of Science and Technology of the People's Republic of China (the National Key Program of Research and Development, 2023YFF1304800 and the Foreign Expert Project, QN2023061006L), the Institute of Zoology, Chinese Academy of Sciences (2023IOZ0104), State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management (SKLA2502 and an open funding grant), the National Natural Science Foundation of China (31970382, 32022009, W2412126 and 32150410358) and the Chinese Academy of Sciences (President's International Fellowship Initiative for Visiting Scientists, 2021PB0022).

Keywords: allopolyploidy | cavefish | genomic differentiation | local adaptation | *Sinocyclocheilus microphthalmus* | subgenome evolution

ABSTRACT

Cave environments impose unique challenges that drive remarkable genetic and phenotypic changes in cave-dwelling organisms. In this study, we investigated the genomic basis of adaptation in the small eye golden-line fish (*Sinocyclocheilus microphthalmus*), an allotetraploid cavefish endemic to Guangxi, China. Using whole-genome resequencing data from 47 individuals across six cave locations, we examined how neutral and selective forces influence diversification. Our analyses uncovered significant population structure indicative of allopatric divergence, along with evidence of locus-specific selection contributing to genomic differentiation. We identified seven single outlier clusters (SOCs), each tied to the divergence of specific populations, underscoring the role of local processes in driving diversity. Genes associated with vision showed relaxed selection, likely reflecting adaptation to darkness, while positive selection on other loci revealed additional functional shifts. Notably, allopolyploidy was found to fuel divergence through subgenome-specific patterns and asymmetric evolution within SOCs and among homoeologs. Taken together, these findings provide valuable insights into mechanisms of cave evolution and illustrate how allotetraploid genomes can facilitate diversification, potentially contributing to speciation in extreme environments.

1 | Introduction

Allopatry is widely recognised as a predominant driver of speciation among populations (Mayr 1942). Under this model, geographically isolated populations gradually become genetically divergent as reduced gene flow allows the accumulation of differences through genetic drift, mutation and natural selection. However, the relative contribution of these evolutionary forces can vary among populations, producing diverse patterns of genomic differentiation. In some cases, divergence is concentrated in specific genomic regions, yielding 'islands of genomic divergence' driven by strong directional selection (Nosil et al. 2009)—a pattern exemplified by studies of *Anopheles gambiae* that identified 'speciation islands' differentiating sympatric forms despite ongoing gene flow (Turner et al. 2005). In other instances, differentiation is more evenly distributed across

Santiago Montero-Mendieta and Yuwei Wang contributed equally to this work.

© 2025 John Wiley & Sons Ltd.

the genome, reflecting drift, variable recombination and background selection (Noor and Bennett 2009), as observed in the repeated adaptation of sticklebacks to freshwater environments (Hohenlohe et al. 2010; Jones et al. 2012). Studying recently diverged populations allows us to observe these processes 'in action' (Wolf and Ellegren 2017), providing critical insights into the early stages of speciation and informing conservation strategies for vulnerable species.

Cave environments, with their isolated nature and strong selective pressures, are ideal natural settings for examining incipient speciation through allopatry (Balogh et al. 2020). Cave-dwelling organisms frequently evolve striking morphological and genetic specialisations to cope with perpetual darkness, limited food and altered ecological interactions (Trontelj 2019). Understanding the relative roles of natural selection and neutral processes in producing these troglomorphic traits is a central question in cave biology. The classic view emphasised the accumulation of neutral loss-of-function mutations under relaxed selection (Wilkens 1988), famously illustrated by the reduction or loss of eyes and pigmentation in many cavefish (Zhang et al. 2025). However, increasing evidence supports a role for positive selection in shaping troglomorphic adaptations (Cartwright et al. 2017)—for example, enhancements of non-visual sensory systems or metabolic shifts that are likely adaptive in cave environments (Moran et al. 2015; Yoshizawa et al. 2015). Variation in effective population size (N_e) among cave and surface populations can further shift the balance between drift and selection: small N_e can magnify drift (e.g., *Astyanax mexicanus*; Culver and Pipan 2019; Fumey et al. 2018), whereas larger cave populations may permit stronger selective effects, as seen in some stygobitic crayfishes (Buhay and Crandall 2005), potentially for heterozygote advantage (Sbordoni et al. 2012).

Endemic to China's southwestern karst region, *Sinocyclocheilus* (Cypriniformes: Cyprinidae) constitutes the world's largest cavefish radiation and is an emerging model for evolutionary research (Mao et al. 2025). Within this genus, the small eye golden-line fish (*Sinocyclocheilus microphthalmus*), a vulnerable (VU) stygobitic species (IUCN 1996), combines three attributes that make it particularly informative to investigate the interplay of evolutionary forces in cave adaptation and diversification. First, unlike many congeners confined to single caves, this species inhabits multiple, isolated cave systems across Guangxi, and maintains relatively large population sizes (Zhao et al. 2021), facilitating comparative genomics. Second, it also exhibits marked habitat-linked phenotypic divergence: in food-rich, partially lit caves, populations tend toward a 'robust' humpback form with larger—albeit still reduced—eyes, whereas those in deep aphotic systems with scarce resources develop a more 'gracile' form with severely regressed eyes (Zhao 2021). Such habitat-phenotype associations enable precise analysis of the genomic basis of troglomorphic adaptations, particularly in the visual system. Importantly, *S. microphthalmus* harbours an allotetraploid genome—resulting from hybridisation and subsequent genome duplication between two divergent parental species. This condition can promote evolutionary innovation through subfunctionalisation and neofunctionalisation (Soltis and Soltis 2012; Van De Peer et al. 2017; Wang et al. 2024), yet its contribution to adaptation in extreme environments such as caves remains poorly understood, particularly in animals.

In this study, we generated and analysed whole-genome resequencing data from 47 *S. microphthalmus* individuals sampled across six distinct cave localities to interrogate the genomic basis of cave adaptation in an allotetraploid vertebrate. We addressed three main questions: (1) What are the relative contributions of neutral processes and natural selection to genomic divergence among populations? (2) How does the allotetraploid genome affect patterns of adaptation across populations? (3) Do visual genes bear signatures of relaxed selection, as predicted by the neutral-mutation hypothesis? To answer these, we analysed nucleotide diversity and linkage disequilibrium, identified highly differentiated loci using linkage disequilibrium network analysis (LDna) and selective-sweep scans, compared variation between subgenomes, examined homoeolog evolution and evidence for subgenome-specific selection, and assessed patterns in visual-system genes. We additionally performed principal component analysis (PCA) and calculated population-genetic statistics within LDna single outlier clusters (SOCs). Our results illuminate the evolutionary forces driving diversification in this vulnerable cavefish and indicate that allopolyploidy can expand evolutionary potential and facilitate adaptation to extreme environments during the early stages of allopatric speciation.

2 | Materials and Methods

2.1 | Sampling, DNA Extraction and Sequencing

We collected 47 *S. microphthalmus* specimens using baited traps between 2001 and 2019 from six localities in Guangxi, China (Figure 1A; Table S1): Fuyuan (FYP; $N=10$), Jiazhuan (ZZP; $N=10$), Jinya (JYP; $N=9$), Luolou (LLP; $N=6$), Sanmenhai (SMHP; $N=3$) and Sicheng (SCP; $N=9$). All fieldwork complied with ethical guidelines and regulations in force at the time and was approved by the Institutional Animal Care and Use Committee of the Institute of Zoology, Chinese Academy of Sciences (IACUC Permit IOZ18002). Although *S. microphthalmus* has been listed as VU by the International Union for Conservation of Nature (IUCN) since 1996, it was not designated a Class II National Key Protected Species in China until 2021—after our sampling—so no additional permits were required. Notably, the sampled populations encompass the species' known morphological range (Zhao 2021): ZZP corresponds to the robust morphotype, whereas FYP, JYP, LLP, SMHP and SCP exhibit the gracile form. DNA extraction, library preparation and sequencing were performed by Novogene Co. Ltd. (Beijing, China). DNA was extracted from fin tissue (DNeasy Blood and Tissue kit, QIAGEN Inc.). Paired-end libraries (from $\geq 1\,\mu g$ DNA/sample) were sequenced on the NovaSeq 6000 platform (Illumina Inc.; mean insert ~350 bp).

2.2 | Reference Genome

Analyses were performed using a chromosome-level reference genome of *Sinocyclocheilus tileihornes*, a closely related species that is also an allotetraploid (Genome Warehouse accession: GWHFQVD00000000.1). The reference genome has a total length of 1.93 gigabase pairs (Gbp) with a scaffold N50 of 35.35 megabase pairs (Mbp). Genome completeness, assessed via BUSCO (v5.4.4; Manni et al. 2021; metazoa_odb9), indicated

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

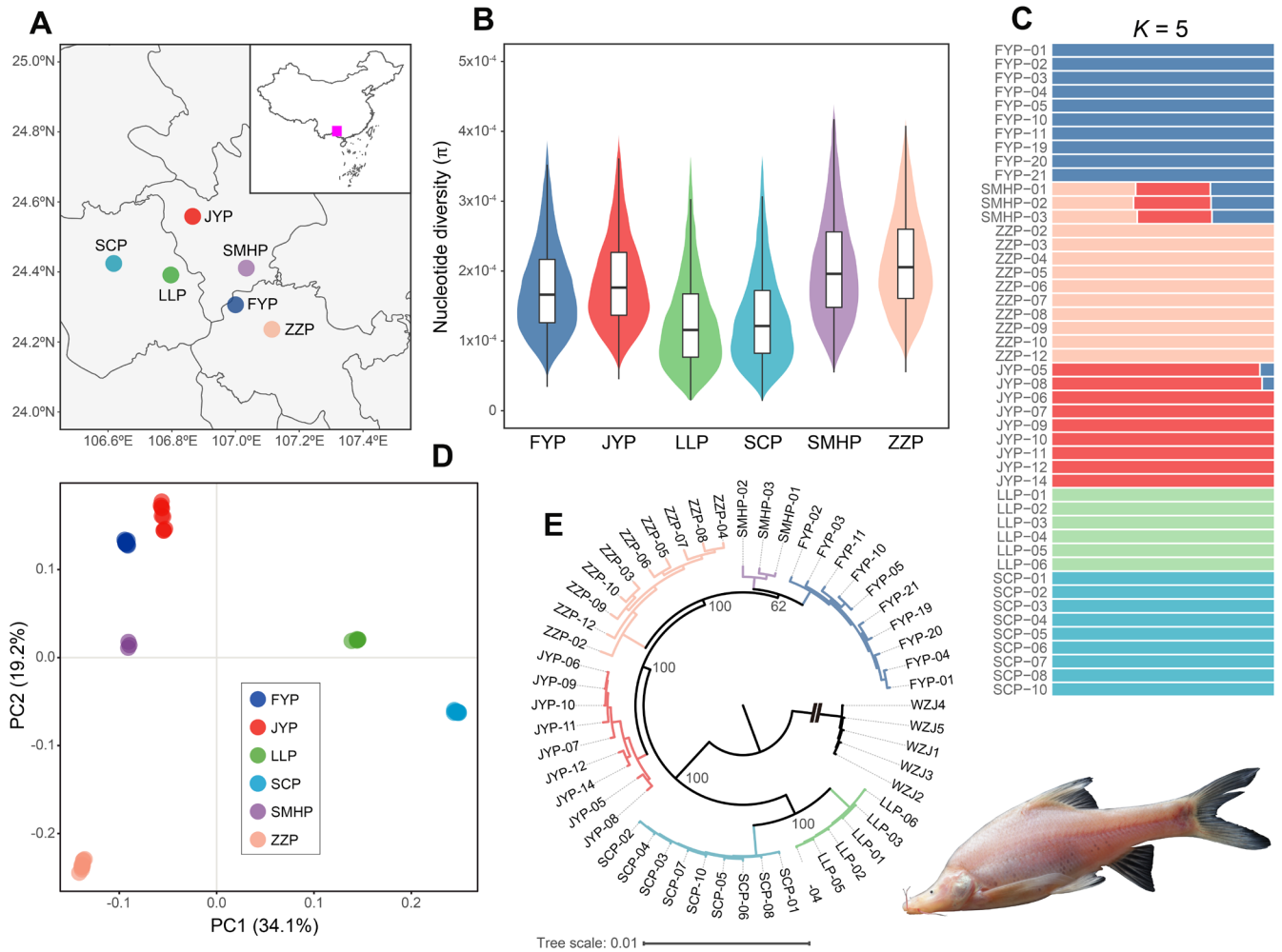

FIGURE 1 | Genetic diversity and population structure of *S. microphthalmus*. (A) Map of the geographic distribution of the six studied populations in Guangxi, China. (B) Nucleotide diversity (π) across six populations, showing significantly lower diversity in LLP and SCP compared to other populations. (C) Population structure inferred from ADMIXTURE analysis at $K = 5$, revealing distinct genetic clusters that largely correspond to sampling localities. (D) Principal component analysis (PCA) of SNP data, with PC1 and PC2 explaining 34.1% and 19.2% of the total variance, respectively, and clearly separating the six populations into distinct clusters. (E) Whole-genome SNP phylogeny illustrating the evolutionary relationships among populations, with SCP and LLP forming a sister clade to the remaining populations. Ultrafast bootstrap support values for major clades are shown next to the nodes. The figure also includes an illustration of an *S. microphthalmus* specimen (Image credit: Yahui Zhao).

that 97.5% of the 978 orthologs searched were found complete: 26.5% as single-copy (S) and 71.0% as duplicated (D). A further 0.9% were found fragmented (F) and 1.6% were missing (M). Based on synteny analysis with the diploid cyprinid *Puntius semifasciolatus*, the genome was resolved into two subgenomes, hereafter referred to as 'heavy' (H) and 'light' (L), each comprising 25 chromosomes. The divergence between the subgenomes, estimated from the median rate of synonymous substitutions (Ks) of syntenic homoeolog pairs, is approximately 17.66 million years ago (Mya).

2.3 | SNP Calling and Annotation

We mapped cleaned reads from all 47 *S. microphthalmus* individuals to the reference genome using BWA-MEM (v0.7.17; Li and Durbin 2009). We then sorted the resulting BAM files and marked PCR duplicates using SAMtools (v1.4; Li et al. 2009) and GATK MarkDuplicatesSpark (v4.1.9; McKenna et al. 2010). We called single nucleotide polymorphisms (SNPs) across all samples using BCFtools (v1.8; Danecek et al. 2021) 'mpileup −Q 20 −q 20', excluding variants within 10 bp of insertions/deletions (indels) ('filter --SnpGap 10'). Subsequently, we used VCFtools (v0.1.16; Danecek et al. 2021) to filter for biallelic SNPs ('--minQ 30 --min-alleles 2 --max-alleles 2 --minDP 2 --maxDP 20 --minGQ 20 --max-missing 0.5 --maf 0.05'). For per-population SNP sets, an additional filter requiring a minor allele count of at least two ('--mac 2') within that population was applied. Finally, we used ANNOVAR (Wang, Li, and Hakonarson 2010) to annotate SNPs by genomic location, subgenome and exonic functional consequence.

2.4 | Phylogenetic Inference and Molecular Dating

To investigate phylogenetic relationships, we first constructed a whole-genome SNP phylogeny. This was inferred using a maximum-likelihood approach in IQ-TREE (v2.0.3; Minh

13652394x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al. 2020) from 2,000,000 randomly subsampled SNPs, with the best-fit model selected via ModelFinder ('-m MFP+ASC') (Kalyaanamoorthy et al. 2017). The input SNP dataset was generated by combining our *S. microphthalmus* data with five *S. tileihornes* outgroups (Genome Sequence Archive accession: CRA023833), followed by variant calling and filtering as described in Section 2.3. Branch support was evaluated with 1000 ultrafast bootstrap (UFBoot) replicates (Hoang et al. 2018).

To estimate divergence times, we separately reconstructed a mitochondrial phylogeny, as the uniparental inheritance of mtDNA avoids complications from homoeologous gene divergence in allotetraploids (Smith and Hahn 2021; Yang et al. 2015). For one individual per cave, we retrieved reads mapping to the *S. tileihornes* mitochondrial genome using SAMtools, and assembled mitogenomes with MitoZ (v1.04; Meng et al. 2019) and MitoS (v2; Bernt et al. 2013). Using 13 protein-coding genes from these mitogenomes and several outgroups (see Supporting Information: Text S1 for full list), we inferred a tree in IQ-TREE with the GTR+F+G4 model and 1000 UFBoot replicates. Molecular dating was then performed using PAML 'MCMCtree' (Yang 2007) with secondary calibration points derived from a comprehensive, time-calibrated *Sinocyclocheilus* genus phylogeny (Figure S1). This broader phylogeny was calibrated using published fossil (Li and Guo 2020) and geological (Li et al. 2008) data. Based on this tree, we applied minimum age constraints of 11.19 Mya at the node separating *S. guilinensis* and *S. jii* from other species, and 5.77 Mya for the most recent common ancestor of *S. anshuiensis* and *S. microphthalmus*. The final dated tree was visualised in FigTree (v1.4.4).

2.5 | Population Genetic Structure and Divergence Estimates

We assessed population genetic structure using ADMIXTURE (v1.3; Alexander et al. 2009)—with the optimal number of genetic clusters (K) (from 1 to 8) determined by the lowest cross-validation error from 10 independent runs per K via CLUMPP (Jakobsson and Rosenberg 2007)—and principal component analysis (PCA; PLINK v1.9; Purcell et al. 2007). To analyse genome-wide divergence, we estimated within-population statistics—nucleotide diversity (π), Tajima's D, observed heterozygous sites (H_O), population recombination rate (ρ), the ratio of nonsynonymous to synonymous polymorphisms (pN/pS)—and between-population statistics—fixation index (F_{ST}), absolute genetic divergence (D_{xy}). These were calculated for combined and individual populations (considering H, L and H + L subgenomes), and in 100 kilobase pair (kbp) non-overlapping sliding windows for π, Tajima's D, ρ, F_{ST} and D_{xy}. We also measured linkage disequilibrium (LD) decay (PopLDdecay v3.40; Zhang et al. 2019) and estimated historical effective population size (N_e) (PSMC; Li and Durbin 2011). To examine the relationship between genetic differentiation and geographic distance, we performed a Mantel test using pairwise F_{ST} values (excluding self-comparisons) and geographic distances (calculated as straight-line measurements between sampling coordinates). Software and parameter details are provided in Supporting Information: Text S2.

2.6 | Scans for Genomic Regions of Differentiation

To identify genomic regions associated with population differentiation, we first performed a linkage disequilibrium network analysis (LDna; Kemppainen et al. 2015), which detects clusters of highly correlated loci (LD-clusters) showing similar population genetic signals. A three-step LDna pipeline (Fang et al. 2020) with parameters from Jin et al. (2022), modified to 'λlim = 2' and '|E|min = 8', was used to extract single outlier clusters (SOCs). For each SOC, we conducted a PCA to determine differentiated (focal) populations. SNPs were classified by subgenome (H, L, H + L), and π, Tajima's D and H_O were calculated for each population. To explore signatures of selection, we then compared diversity statistics between focal and non-focal populations using Wilcoxon rank-sum tests. Additionally, pairwise F_{ST} and D_{xy} were estimated for all population pairs within each SOC and subgenome.

In parallel, we scanned the genome for signatures of recent positive selection using SweeD (v4.0.0; Pavlidis et al. 2013) and RAiSD (v2.9; Alachiotis and Pavlidis 2018) on combined and per-population datasets. Putative selective sweeps were defined as the top 5% shared outliers located within 400 bp of each other across both methods (details in Supporting Information: Text S2). Finally, Gene Ontology (GO) enrichment analysis (g:Profiler2; Kolberg et al. 2020) was performed on the gene sets identified from both the SOC and the sweep scans.

2.7 | Assessment of Genetic Variation Between Subgenomes

We established one-to-one syntenic blocks between the H and L subgenomes (MUMmer v4; Marçais et al. 2018), and one-to-one homoeologous genes within these blocks using all-vs-all BLASTp (E-value $\leq 1\times10^{-5}$; Altschul et al. 1990) and MCScanX (Wang et al. 2012). These homoeologs were used to calculate SNP density (SNP/kbp) per subgenome (H and L) in syntenic blocks, homoeologs and their coding sequences (CDS), including densities of synonymous (synSNPs) and nonsynonymous (nsSNPs) within CDS. Differences between subgenomes were tested using Wilcoxon rank-sum tests.

2.8 | Selection Direction Analysis

To explore evolutionary forces driving subgenome differentiation, we identified homoeologs under putatively similar selective pressures by comparing each pair's nsSNP to the overall average nsSNP density across all homoeologs. Homoeolog pairs with both copies exhibiting nsSNP densities $<0.5\times$ or $>2\times$ this average were classified as under purifying or relaxed selection, respectively. We investigated asymmetric evolution between H and L subgenome copies by calculating the $\log_2$ ratio of their nsSNP densities (adding a pseudo-count of 1×10^{-7} to handle zero values: $\log_2[(\text{nsSNP\_density\_H} + 1\times10^{-7})/(\text{nsSNP\_density\_L} + 1\times10^{-7})]$). We also tested for significant differences in the number of nsSNPs and nonsynonymous sites (identified using KaKs_Calculator v2.0; Wang, Zhang, et al. 2010) between subgenome copies using

13652294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fisher's exact test. Homoeologs with a $|\log_2 \text{ratio}| > 1$ and a Benjamini–Hochberg (BH) corrected q-value < 0.05 from Fisher's test were considered asymmetrically evolved. These analyses were performed for all populations combined and for each population separately.

2.9 | Visual Genes and Evolutionary Rate Shift Analyses

To assess selection pressures on visual genes, we used a candidate gene list from Yang et al. (2016) and compared their nsSNP density to the genome-wide average using a Wilcoxon rank-sum test. We also investigated evolutionary rate shifts in *S. microphthalmus* by comparing its genes to their human orthologs, using the latter as a baseline for conserved vertebrate evolutionary rates (Ka/Ks data from The Chimpanzee Sequencing and Analysis Consortium 2005). After excluding genes with Ks = 0, we ranked human genes by their Ka/Ks ratio and classified them into three equal-sized categories: 'slow' (lowest third), 'medium' (middle third) and 'fast' (highest third). *S. microphthalmus* genes were similarly grouped by nsSNP density into 'low', 'medium' and 'high', with genes lacking nsSNPs included in the 'low' category. We then identified genes with contrasting rate categories between the species (i.e., those slow-evolving in humans but with high nsSNP density in *S. microphthalmus* and vice versa) for GO enrichment analysis (g:Profiler2). As a final step to assess asymmetric evolution, we focused on the homoeolog pairs within these contrasting sets. For each pair, the homoeolog with the rate that defined its inclusion in a set (e.g., the 'slow' copy in the 'Human Fast, *S. microphthalmus* Slow' group) was termed the 'diverged copy'. We then measured the rate of its corresponding 'partner copy' to test whether it had instead retained the conserved, human-like evolutionary rate.

3 | Results

3.1 | Genetic Diversity and Population Structure

Our study focused on six cave populations of *S. microphthalmus* from Guangxi, China, spanning the known morphological and ecological spectrum of the species (Figure 1A; Table S1). Based on known phenotypes (Zhao 2021), the Jiazhuan (ZZP) population corresponds to the robust morphotype from a cave connected to the surface via sinkholes, a habitat that is partially lit and food rich. The remaining five populations—Fuyuan (FYP), Jinya (JYP), Luolou (LLP), Sanmenhai (SMHP) and Sicheng (SCP)—belong to the gracile morphotype from deep, aphotic cave systems.

Whole-genome resequencing of 47 *S. microphthalmus* individuals (average depth 12.46×; 98.63% mapping rate; Table S1) yielded 1,241,186 high-quality SNPs (average nucleotide diversity: $\pi = 2.30 \times 10^{-4}$), primarily located in intergenic (51.38%) and intronic (38.12%) regions, with 5.96% exonic (Table S2). Per-population SNP counts ranged from 641,919 (LLP) to 1,443,887 (ZZP). Pairwise chi-squared (χ^2) tests revealed significant differences in SNP proportions across genomic regions among populations ($p < 0.05$; Supporting Information: Text S3). These deviations were most pronounced in the gracile populations LLP and SCP, particularly in exonic SNP categories; these two populations also exhibited significantly lower nucleotide diversity (mean $\pi = 1.44 \times 10^{-4}$ vs. 2.06×10^{-4} in others; $p < 0.001$; Figure 1B), Tajima's D, observed heterozygosity (H_O) and population recombination rate (ρ) (all $p < 0.05$). Conversely, their pN/pS ratios were higher (1.06–1.07 vs. 1.02–1.03), and their historical effective population size (N_e) trajectories remained consistently smaller while other populations experienced a period of growth and decline (Figure S2). Separately, linkage disequilibrium (LD) decayed more slowly in LLP and SMHP, though the result for the latter may be influenced by its smaller sample size ($N = 3$) (Figure S3).

Population structure analyses clearly defined the relationships among populations. ADMIXTURE revealed five distinct genetic clusters ($K = 5$) that largely corresponded to sampling localities (Figure 1C). PCA corroborated these findings, grouping the six populations into distinct clusters and explaining 53.3% of total variance with its first two principal components (PC1 and PC2). PC1 primarily distinguished the highly divergent LLP and SCP populations, while PC2 further separated the robust ZZP population (Figure 1D). Notably, the minor introgression signal observed for the SMHP population in the full ADMIXTURE analysis was absent upon subsampling ($N = 3$ per population; Figure S4), suggesting it was a computational artefact.

3.2 | Genome-Wide Differentiation

The whole-genome SNP phylogeny placed SCP and LLP as a sister clade to the remaining populations, within which JYP diverged first, followed by ZZP, then a SMHP–FYP clade (Figure 1E). The mitochondrial phylogeny differed, placing SCP as the earliest diverging lineage and JYP sister to SMHP/FYP (Figure S5). Divergence time estimates from the mitochondrial phylogeny suggest population splits occurred within the past ~240,000 years: SCP (0.15–0.24 Mya), LLP (0.14–0.22 Mya), ZZP (0.13–0.20 Mya), JYP (0.08–0.15 Mya), with the most recent divergence between SMHP and FYP (0.02–0.06 Mya). In both trees, SMHP and FYP formed the closest pair (0.02–0.06 Mya). This pairing appeared in both trees, though PCA showed FYP and JYP as nearly overlapping clusters (Figure 1D).

Substantial genetic differentiation among populations was observed (Table S3), with pairwise F_{ST} ranging from 0.116 (SMHP–ZZP) to 0.372 (SMHP–SCP) and D_{xy} from 0.282 (LLP–SCP) to 0.385 (SCP–ZZP). LLP and SCP showed significantly higher mean pairwise F_{ST} (0.318 and 0.343, respectively) than other populations (0.216–0.231; Wilcoxon, $p < 0.01$), a pattern robust to downsampling ($N = 3$ per population; Pearson $R = 0.98$, $p < 0.001$; Figure S6). In contrast, mean D_{xy} showed no clear trend (0.334–0.359). The high F_{ST} for LLP and SCP aligned with their lower π, H_O and ρ, and the two were also highly differentiated from each other ($F_{ST} = 0.323$). A neighbour-joining tree based on F_{ST} distances further highlighted their distinctiveness, placing them on the longest branches (Figure S7), with relationships among other populations generally congruent with the mitochondrial phylogeny. No significant correlation was found between genetic and geographic distance (Mantel test $r = 0.35$, $p = 0.11$) (Figure S8).

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

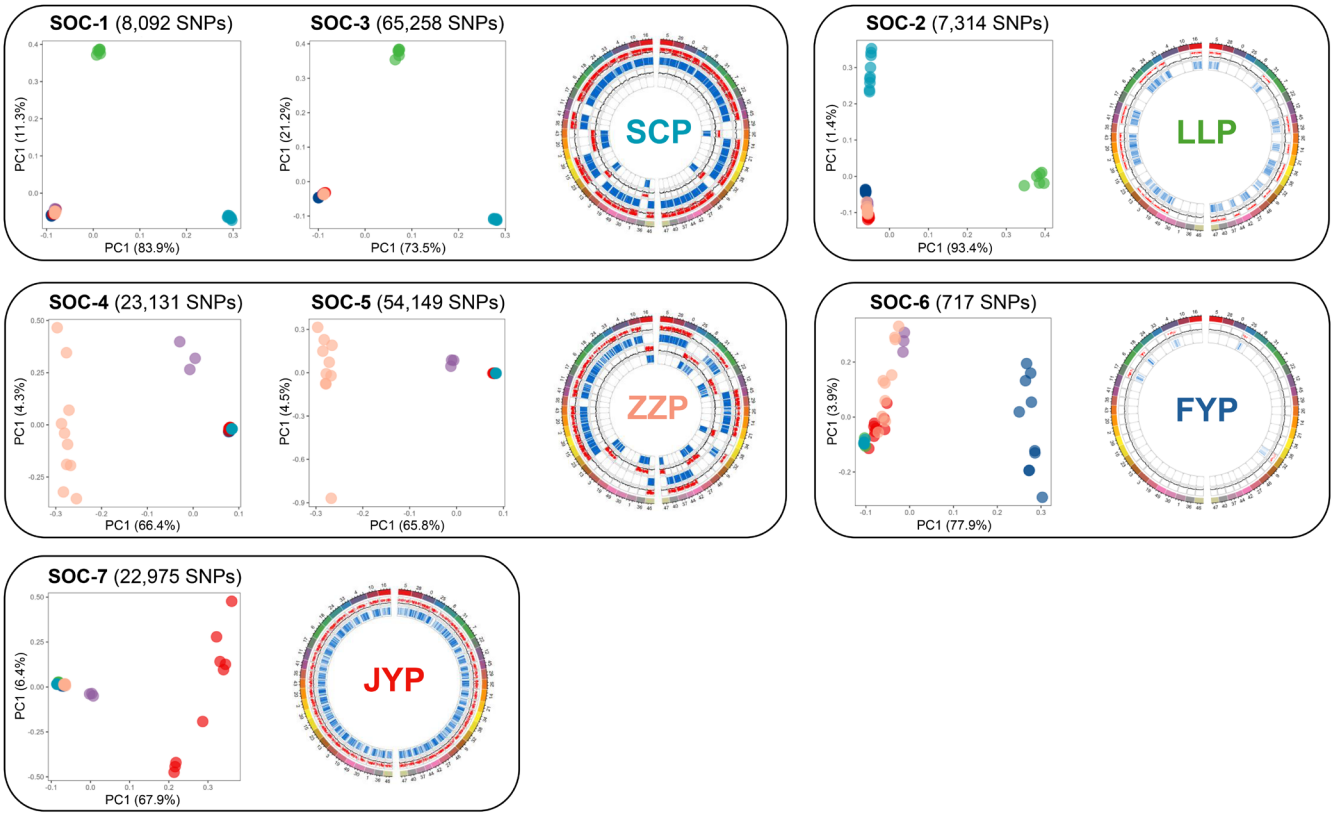

FIGURE 2 | Principal component analysis (PCA) and genomic distribution of SNPs within each single outlier cluster (SOC) identified through linkage disequilibrium network analysis (LDna). Each panel displays PCA results for specific SOCs, highlighting populations with distinct genetic differentiation. The right side of each panel shows the genomic distribution of SOC SNPs in the most differentiated population, with the outer circle representing chromosomes, the middle circle (red) displaying F_{ST} for SOC SNPs (black line indicating genome-wide F_{ST}), and the inner circle (blue) showing SNP density. These results highlight how distinct genomic regions contribute to genetic differentiation within each cave environment.

3.3 | Signatures of Selection

Linkage disequilibrium network analysis (LDna) identified seven single outlier clusters (SOCs) comprising 181,636 SNPs, representing genomic regions potentially shaped by local selection. These SOCs varied in genomic coverage, from SOC-6 (717 SNPs across five chromosomes) to SOC-3 (65,258 SNPs across 50 chromosomes) (Table S4; Figure S9), and showed significant differences in their SNP distributions across genomic regions (χ^2 tests). PCA based on SNPs within each SOC distinguished five of the six populations, aligning with genome-wide findings (Figure 2; Section 3.1)—a pattern supported by elevated F_{ST} and D_{xy} between focal and non-focal populations within each SOC (Table S3). Divergent (focal) populations within these clusters showed significantly higher π and Tajima's D (Wilcoxon $p < 0.05$), while H_O did not differ (Table S4). Functional annotation of genes within SOCs revealed notable enrichment for binding (e.g., protein, calcium ion, ATP) and catalytic activities (e.g., ATPase), as well as biological processes like molecular transport and signal transduction (Table S5).

Genome scans for recent positive selection using SweeD and RAiSD identified 15 selective sweeps in the combined dataset (spanning 1.86 Mbp and 53 genes across 13 chromosomes; Figure S10; Table S6). These sweeps contained genes possibly linked to cave-related traits (e.g., mtnr1c, npas2, opn4xb, CYP26C1, TGM2), and were enriched for GO terms related to transferase activity and peptide cross-linking (Table S7). Population-specific scans detected a greater number of sweeps (from 20 in LLP to 43 in ZZP). Notably, the highly differentiated LLP and SCP populations had among the fewest such sweeps. Only FYP and SMHP showed significant GO enrichments in their population-specific scans—FYP for kinetochore attachment and protein polymerisation; SMHP for metabolic, biosynthetic and vitamin A-related processes. No genes were shared between these population-specific and the combined sweep sets. More broadly, no SNP overlap was found between any SOC and sweep region, evidencing these analytical approaches capture distinct evolutionary signatures.

3.4 | Subgenome Evolution

Analysis of the two constituent subgenomes of *S. microphthalmus*, designated H ('heavy') and L ('light'), uncovered distinct patterns of variation. Specifically, the H subgenome (9.06×10^8 bp) harboured slightly more polymorphic sites (641,634 SNPs; 51.70%) than the L subgenome (8.83×10^8 bp; 599,552 SNPs; 48.30%) (Table S8). At the same time, π was markedly higher in the H subgenome across all populations (Wilcoxon $p < 0.05$; Figure S11; Table S2), as was Tajima's D specifically in the SMHP population ($p < 0.001$). Regarding selection on coding regions, the pN/pS ratio showed contrasting

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

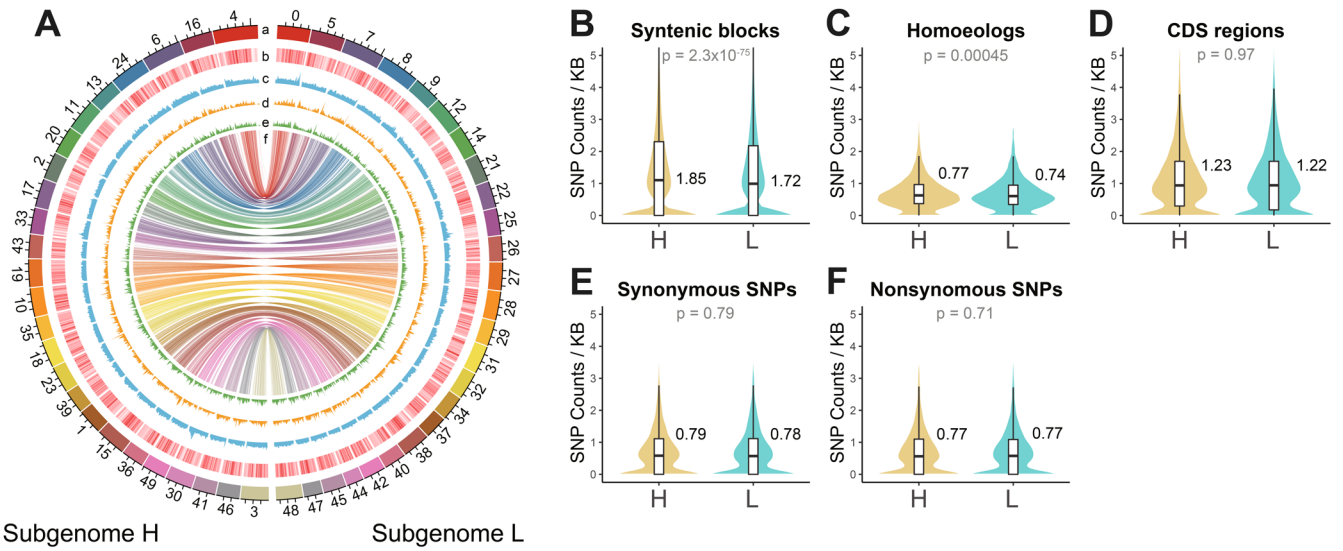

FIGURE 3 | Genome-wide distribution and SNP density comparisons between H and L subgenomes. (A) Circular representation of the *S. microphthalmus* genome. From outer to inner circles: (a) chromosome labels, (b) gene locations (red), (c) SNP distribution (blue), (d) synonymous SNPs (orange), (e) nonsynonymous SNPs (green) and (f) syntenic gene pairs. (B–F) Comparisons of SNP density (SNP/kbp) between H and L subgenomes for: (B) syntenic blocks, (C) homoeologs, (D) CDS regions, (E) synonymous SNPs, and (F) nonsynonymous SNPs. SNP density was significantly higher in the H subgenome for syntenic blocks and homoeologs ($p < 0.001$, Wilcoxon rank-sum test), but not for CDS regions, synonymous SNPs, or nonsynonymous SNPs.

patterns: it was elevated in the H subgenome for LLP and SCP, but higher in the L subgenome for other populations. However, no clear overall differences were observed between subgenomes for Tajima's D, H_O, ρ, or genome-wide F_{ST} and D_{xy} (Tables S2 and S3).

At a finer scale, SNP distributions differed substantially across the genome between subgenomes (pairwise χ^2, $p < 0.001$; Table S2), with intergenic and intronic regions showing the largest deviations. Notably, LLP and SCP exhibited a pronounced excess of SNPs in exons (mean standardised residual [SR], $p < 0.01$) and for nsSNPs ($p < 0.001$) relative to other populations. This subgenome bias was also evident within most SOCs, where the H subgenome contained more SNPs (~51%–75%), except for SOCs 3 and 4, in which the L subgenome predominated (Table S4). Separately, analysis of subgenome-specific diversity within SOCs revealed significant results in certain focal populations. In LLP, Tajima's D was elevated in the H subgenome in SOC-1 and SOC-3 ($p < 0.001$ and $p < 0.05$, respectively), while in SCP, H_O showed a greater value in the L copy in SOC-1 ($p < 0.01$). Additionally, the H subgenome in SOC-6 (which differentiates FYP) displayed increased Tajima's D and H_O ($p < 0.01$ and $p < 0.05$), whereas in SOC-7 (JYP) had a larger π ($p < 0.05$). No consistent subgenome differences in F_{ST} or D_{xy} were found between focal and non-focal populations (Table S3).

3.5 | Homoeolog Analysis

A total of 13,705 homoeolog pairs were identified within 232,505 syntenic blocks. Overall, SNP density (SNP/kbp) was significantly higher in the H subgenome for syntenic blocks (H = 1.85 vs. L = 1.72, $p < 0.001$) and homoeologs (H = 0.77 vs. L = 0.74, $p < 0.001$), but not for their CDS regions ($p = 0.97$), a pattern robust to coverage filtering (Figure 3; Table S2; Figure S12). This trend of higher H-copy variation was generally consistent across individual populations, though CDS differences were limited to JYP, LLP and SCP (Figures S13–S17).

Based on nsSNP density (Figure 4A), 545 homoeolog pairs were classified as under purifying selection (< 0.5× average) and 172 as under relaxed selection (> 2× average). Homoeologs under purifying selection showed enrichment for GO terms related to essential functions like protein/ATP binding, myosin complex, and transport processes (Figure 4B; Table S9). In contrast, those under relaxed selection were remarkably enriched for visual perception and phototransduction, among other functions (Figure 4C; Table S9). These patterns varied among populations; notably, the gracile populations LLP and SCP had a significantly higher percentage of homoeologs under purifying selection (~90% and ~85%, respectively; Fisher's $p < 0.001$) compared to others (~76%–80%) (Figure S18). While GO enrichments for purifying selection were consistent across all populations, those for relaxed selection were largely population-specific (Table S10), with shared terms including vision-related functions (e.g., visual perception in SMHP/ZZP, phototransduction in LLP/ZZP) and developmental processes. Immune-related functions under purifying selection were exclusively enriched in LLP and SCP.

Investigation of asymmetric evolution revealed that 1086 homoeologs had a $|\log_2$ ratio$|$ of nsSNP densities > 1 (Figure S19). Fisher's exact tests identified three significantly asymmetric homoeolog pairs (q-value < 0.05; Table S11): two associated with nervous system genes (CACNA2D2, PCDHA3 and PCDHAC2) had elevated nsSNP density in the H subgenome, while the third, an immune-response gene (FCGBP), displayed a greater density in the L copy. Population-specific analyses confirmed CACNA2D2 consistently possessed an excess of nsSNP density in the H subgenome of three gracile populations (JYP, LLP

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

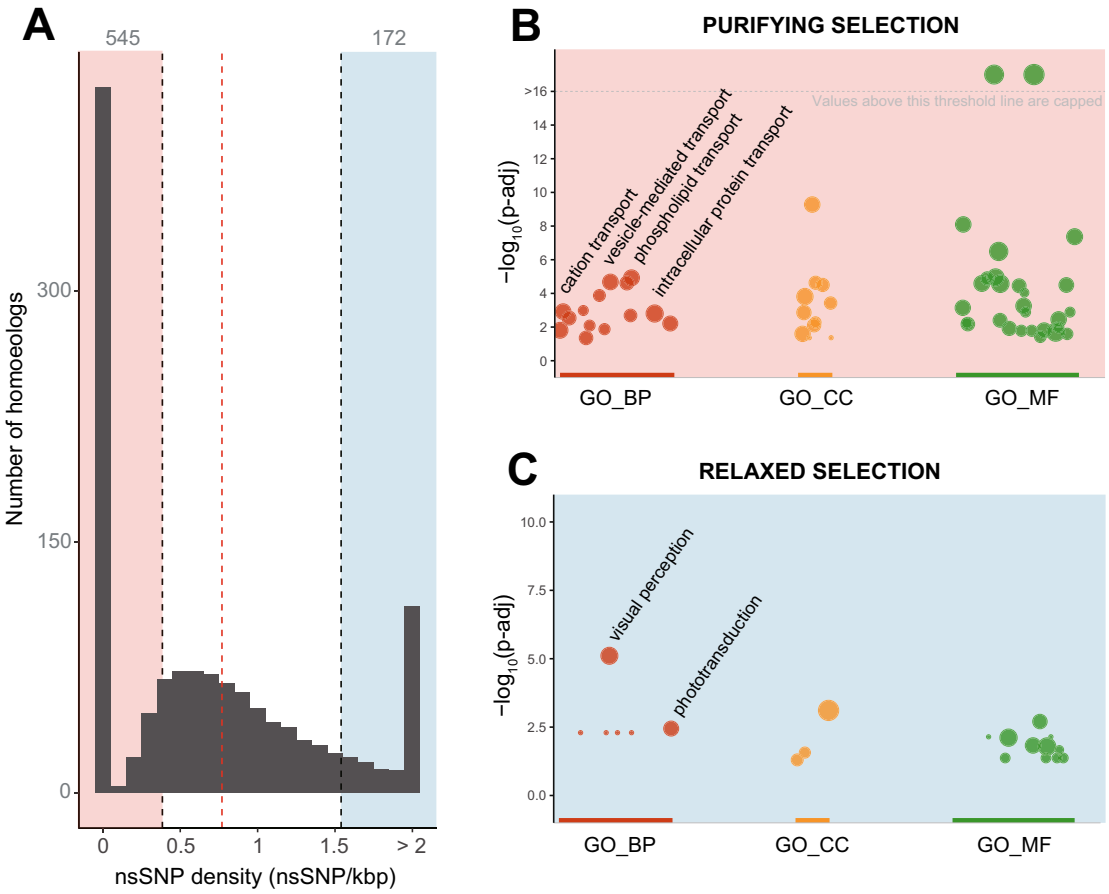

FIGURE 4 | Classification of homoeologs under purifying or relaxed selection based on nonsynonymous SNP (nsSNP) density and associated GO enrichment. (A) Distribution of nsSNP density among homoeologs. The red dashed line represents the overall average nsSNP density, while black dashed lines indicate thresholds for purifying selection (nsSNP density $<0.5\times$ average) and relaxed selection (nsSNP density $>2\times$ average). (B) GO enrichment analysis of homoeologs under purifying selection, showing significantly enriched biological processes (GO_BP), cellular components (GO_CC), and molecular functions (GO_MF). (C) GO enrichment analysis of homoeologs under relaxed selection, with top enriched terms related to visual perception, transcription factor binding, and lipid binding.

and SMHP), with an exceptionally strong signal in LLP ($\log_2$ ratio = 16.55, $p < 0.001$). SCP was the only population lacking significant asymmetrically evolving homoeologs (Figure S20; Table S12).

3.6 | Visual Genes and Evolutionary Rate Shifts

Homoeologs related to visual genes exhibited a higher average nsSNP density (1.01 nsSNPs/kbp) compared to the genome-wide average for all homoeologs (0.77). While no overall significant differences in visual gene nsSNP density were detected between H and L subgenomes across populations (Wilcoxon $p > 0.05$), the SCP population showed a striking pattern: its L subgenome visual genes had markedly higher nsSNP density (1.12) than both its H subgenome counterpart (0.56) and the respective genome-wide averages. LLP and SCP were the only populations where the L copy of visual genes had higher nsSNP density than the H copy.

A comparison of evolutionary rates between *S. microphthalmus* and human orthologs identified 1195 genes evolving rapidly in humans (high Ka/Ks) but slowly in *S. microphthalmus* (low nsSNP density) (i.e., 'Human Fast, *S. microphthalmus* Slow') and 281 genes with the opposite pattern (i.e., 'Human Slow, *S. microphthalmus* Fast'). The former set was mainly enriched for transcription/RNA processing, oxidoreductase activity and molecular transport (Table S13), and included 228 homoeolog pairs where both copies evolved slowly in *S. microphthalmus*—linked to processes such as DNA replication and cytoskeletal organisation. Conversely, the latter set showed enrichment for transcriptional regulation, proteolytic activity and ion transport. Notably, this set only had 13 homoeolog pairs where both copies evolved rapidly in *S. microphthalmus*, and these were overrepresented in phototransduction and oxidative stress response functions.

To assess asymmetric evolution, cases where only one homoeolog per pair had diverged from the human ortholog rate were examined. Applying this criterion to the 'Human Fast, *S. microphthalmus* Slow' set identified 340 such homoeolog pairs. Within these pairs, the 'diverged copy' (the one that evolved slowly; median nsSNP density: 0.11) was accompanied by a 'partner copy' that had retained the fast, human-like evolutionary rate (median nsSNP density: 1.04) (Figure 5). Similarly, in the 'Human Slow, *S. microphthalmus* Fast' set, this resulted in 173 such pairs where the diverged copy (which evolved rapidly; median nsSNP density: 2.03) had a partner that preserved the slow, conserved rate (0.28).

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

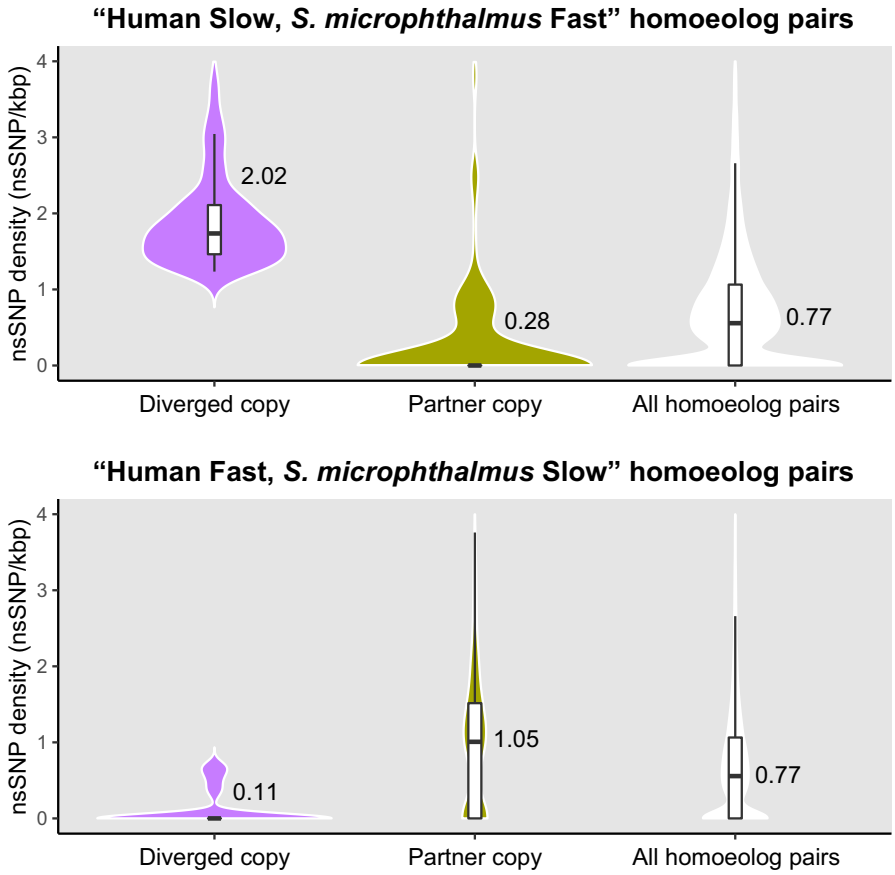

FIGURE 5 | Asymmetric evolutionary rate shifts in homoeologs compared to human orthologs. Violin plots show the distribution of nonsynonymous SNP (nsSNP) density in homoeolog pairs where only one copy has diverged from the conserved human evolutionary rate. The upper panel represents pairs from the 'Human Fast, *S. microphthalmus* Slow' set; the lower panel represents pairs from the 'Human Slow, *S. microphthalmus* Fast' set. In each panel, the 'diverged copy' (purple) is the homoeolog that shifted its rate, while the 'partner copy' (olive green) is the one that was subsequently measured and found to have retained the human-like rate. All homoeolog pairs (white) are shown for reference.

4 | Discussion

In this study, we investigated the genomic basis of adaptation and diversification in the allotetraploid cavefish *S. microphthalmus*, focusing on populations from six caves in Guangxi, China. By integrating population genomics, phylogenetics and subgenome evolution analyses, we sought to understand the interplay of evolutionary forces shaping the genome of this polyploid species in extreme environments. Our results reveal several key findings. First, populations show substantial genetic differentiation (pairwise F_{ST}: 0.116–0.372), consistent with the isolated nature of cave habitats. Second, we identified seven single outlier clusters (SOCs)—loci groups with elevated linkage disequilibrium (LD)—pointing to balancing selection and local adaptation. Third, despite similar genome-wide differentiation between the H and L subgenomes, the H copy consistently carried more SNPs and higher nucleotide diversity (π)—underlining the asymmetry often seen in polyploid genomes. Fourth, at the homoeolog level, most gene pairs remain under strong purifying selection, but a subset displayed relaxed selection on visual functions. Finally, we found evidence for evolutionary rate shifts in several genes, some evolving faster in *S. microphthalmus* than in their human orthologs, possibly reflecting cave adaptation. Taken together, these findings illustrate how allopolyploidy, selection and neutral processes interact to drive diversification and provide evolutionary pathways for incipient speciation.

4.1 | Population Structure and Divergence: Allopatry and Recent Isolation

Cavefishes, often isolated in subterranean habitats, represent striking examples of allopatric divergence (Jeffery 2009; Juan et al. 2010). Our findings in *S. microphthalmus* strongly support this view: in addition to high F_{ST} values, multiple analyses—including ADMIXTURE, PCA and phylogenetic inference from both mitochondrial and nuclear data—confirm that the six sampled populations are distinct evolutionary lineages. Divergence-time estimates place their separation within the past 240,000 years, a period marked by climatic oscillations that could have reshaped cave connectivity and promoted isolation (Culver and Pipan 2019). However, the lack of correlation between genetic and geographic distance in the Mantel test suggests that factors beyond isolation-by-distance have influenced the observed genetic patterns.

Two gracile morphotype populations from deep-cave systems, Luolou (LLP) and Sicheng (SCP), emerged as particularly distinct, characterised by lower π, Tajima's *D*, observed

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

heterozygosity (H_O) and population recombination rates (ρ). These signatures point to long-term reductions in effective population size (N_e), likely through founder events or bottlenecks (Frankham et al. 2002; Nei 2005), a pattern further supported by PSMC trajectories. However, the relatively recent divergence of this population pair from others (0.14–0.25 Mya), along with the lack of consistent D_{xy} signals, suggests that genetic drift alone is insufficient to explain their strong differentiation. Instead, an excess of nonsynonymous SNPs (nsSNPs) in exonic regions, together with elevated pN/pS ratios, indicate relaxed or even positive selection acting on the coding sequences of LLP and SCP (Yang and Bielawski 2000). Additionally, we found that LD decayed more slowly in LLP and Sanmenhai (SMHP) than in other populations. While slower LD decay in SMHP could stem from a small sample size ($N = 3$), the pattern in LLP may reflect reduced recombination rates or increased inbreeding, which can preserve linkage disequilibrium over longer genomic distances (Slatkin 2008).

4.2 | The Role of Selection in Shaping Genomic Divergence

The discovery of seven SOCs, each primarily associated with one or two populations, underscores how distinct genomic regions contribute to genetic differentiation across cave habitats. These SOCs varied widely in SNP content, from a few hundred SNPs in SOC-6 to tens of thousands in SOC-3, but we found no clear link between SNP count and chromosomal coverage. This decoupling suggests that SOC size and location are shaped not solely by genome-wide patterns of variation but also by heterogeneous selective pressures (Charlesworth 2006; Cruickshank and Hahn 2014). Functionally, SOCs consistently harboured genes enriched for protein, calcium ion and ATP binding, as well as molecular transport and signal transduction—processes likely critical for responding to cave-specific stressors like hypoxia or nutrient limitation (Gundemir et al. 2012). Significantly, this enrichment for core cellular functions overlaps with both highly conserved and rapidly evolving gene sets (Sections 3.5 and 3.6), suggesting that local adaptation in this system frequently proceeds through fine-tuning of essential pathways.

The specific selective forces at play appear to vary among these SOCs. Elevated π, Tajima's D and H_O in the differentiated populations within SOCs 4, 5, 6 and 7 point to balancing or diversifying selection maintaining genetic variation. This is particularly evident in the robust morphotype population, Jiazhuan (ZZP), where the primary regions of genomic differentiation (SOCs 4 and 5) show these high diversity metrics coinciding with GO enrichment for synaptic, GABA-signalling, and cell-migration functions (Table S7)—a combination in line with behavioural and developmental flexibility in its complex, partially lit habitat. A similar pattern is seen in SOC-7, which differentiates the gracile Fuyuan (JYP) population, although here ion transport and nutrient-sensing functions are overrepresented, potentially reflecting adaptation to resource scarcity in the deep-cave system. In contrast, SOCs 1 and 3 (SCP and LLP) show comparable levels of diversity between focal and non-focal populations: Tajima's D hovers near zero and H_O remains similar, implying relaxed constraint at these loci (Lahti et al. 2009). SOC-2 (LLP), with its negative Tajima's D, likely reflects demographic history rather than strong selection.

Notably, none of the SOCs overlapped with selective sweep regions, indicating that they capture different evolutionary processes: SOCs trace genomic areas shaped by longer-term balancing selection or local adaptation, whereas sweeps pinpoint loci under more recent, rapid positive selection (Montero-Mendieta et al. 2019). Our genome scans identified 15 sweeps across the species and up to 43 in individual populations. Candidate genes within these regions were linked to circadian rhythm, eye development, stress response and metabolism (Cvekl and Wang 2009; Kitahashi et al. 2021; Mack et al. 2021; Sajovic et al. 2022), highlighting their importance in cave-specific traits. ZZP held the highest number of sweeps, aligning with the expected dynamism of its sinkhole-connected habitat, which forms a direct interface between the cave and surface and allows for selection on both subterranean and epigean traits. By comparison, the limited sweep signal in the highly divergent LLP and SCP populations shows that recent positive selection contributed little to their genomic divergence.

4.3 | Subgenome Dynamics and Asymmetric Evolution

The allotetraploid genome of *S. microphthalmus* adds a fascinating layer of evolutionary complexity. On one hand, it provides unique opportunities for adaptation, as functional redundancy from duplicated gene sets can buffer against deleterious mutations and provide raw material for evolutionary innovation (Soltis and Soltis 1999; Van De Peer et al. 2017; Li and Guo 2020). On the other hand, it presents analytical challenges, such as accurately disentangling the evolutionary histories of the two subgenomes. Our analyses revealed that while most genome-wide diversity and differentiation metrics were generally similar between the H and L subgenomes, π was consistently higher in the H copy. Given that SNP distribution differences were most pronounced in non-coding regions where purifying selection is typically weaker (Lynch 2007; Ronfort 1999), this pattern likely reflects a general relaxation of selective constraints on the H subgenome. Alternative explanations, such as differences in mutation rates or demographic history between subgenomes, are less likely, as these would affect coding and non-coding regions more uniformly (Otto and Whitton 2000; Wolfe 2001).

The pN/pS ratio, however, varied across populations and between subgenomes, evidencing population-specific selection in coding sequences: it was higher in the H subgenome for LLP and SCP, but higher in the L for others. This disparity suggests that differential selective regimes on the two subgenomes may underlie the marked divergence of this population pair (Edger and Pires 2009). In LLP, elevated Tajima's D in the H copy of SOCs 1 and 3, alongside high pN/pS ratios and no overlap with sweeps, is indicative of relaxed purifying selection. In SCP, by contrast, higher H_O was found in the L copy of SOC-1, suggesting balancing selection. Together, these results point to a complex partitioning of evolutionary roles, where the H subgenome may contribute more strongly to broad-scale

13652294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adaptive divergence, while the L subgenome can either retain genetic diversity at key loci (e.g., SOC-1 in SCP) or act as a source of specific gene-level variation, with the latter role being supported by patterns of asymmetric homoeolog evolution (see below).

At the homoeolog level, most duplicated gene pairs were found to evolve under purifying selection, consistent with the preservation of essential functions (Blanc and Wolfe 2004). However, a subset of homoeologs showed signatures of relaxed selection, an outcome further explored in the context of visual gene evolution (Section 4.4). Homoeologs under the purifying regime were enriched for core cellular processes such as protein/ATP binding and molecular transport. Interestingly, and in apparent contrast to their genome-wide signal, LLP and SCP exhibited a greater proportion of such genes than other populations. While their smaller N_e might be expected to reduce the efficacy of purifying selection, this supports a compelling 'two-speed' evolutionary scenario: their genomic background may be evolving under relaxed constraint, but local pressures (e.g., reduced parasite diversity, cave-specific pathogens) may simultaneously intensify selection on key functional categories (Peuß et al. 2020), as evidenced by the exclusive enrichment for immune-related terms among homoeologs under purifying selection in these two populations.

Patterns of asymmetric evolution between homoeologs provide a more in-depth view of subgenome partitioning. Although most homoeologs showed balanced nsSNP densities, a subset exhibited marked asymmetry. Among the statistically significant cases, a clear trend emerged along functional lines: two pairs associated with nervous system genes (CACNA2D2, PCDHA3 and PCDHAC2) showed higher nsSNP densities in the H subgenome, while a third, the immune-related gene FCGB, had elevated density in its L copy. The asymmetric evolution of the CACNA2D2 homoeolog pair, a calcium channel gene, is particularly noteworthy as it was independently detected in three gracile morphotype populations (JYP, LLP and SMHP). This gene-specific pattern aligns with our broader finding that 'calcium ion binding' was a common functional category enriched across SOCs (Section 3.3), and strongly suggests that the H subgenome is a primary locus for adaptation in neurological pathways—potentially contributing to eye degeneration (Orrenius et al. 2003; Jeffery 2009). In contrast, the pattern for FCGBP points to a more prominent adaptive role for the L subgenome in immune function. These results bolster the idea that the two subgenomes promote divergence and local adaptation in distinct ways, though functional validation will be necessary to confirm their phenotypic relevance.

4.4 | Visual Gene Evolution, Rate Shifts and Allopolyploidy's Influence

Multiple lines of evidence from our analyses converge to support pervasive relaxed selection shaping visual gene evolution in *S. microphthalmus*, a process common in cave-adapted species (Jeffery 2001; Protas et al. 2006). First, visual gene homoeologs as a group exhibited a higher average nsSNP density than the genome-wide background (Section 3.6), indicating diminished purifying selection (Calderoni et al. 2016). Second, homoeologs classified as being under relaxed selection (Section 3.5) were notably enriched for functions in visual perception and phototransduction. Third, in our comparison with human orthologs, the 13 homoeolog pairs that evolved rapidly in the cavefish were also enriched for these same functions (Section 3.6). Jointly, this consistent accumulation of non-synonymous changes, especially the observation that it occurs in both homoeolog copies of key visual pathway genes, implies that the functional redundancy from allopolyploidy may actually accelerate the process of neutral decay in the visual system (Culver et al. 2023; Zhao et al. 2022). This effect also appears to be subgenome-specific in SCP, a gracile morphotype with severely regressed eyes, where the L copies of visual genes exhibit markedly higher nsSNP densities, suggesting the process of eye degeneration may follow distinct evolutionary paths across subgenomes.

Broader evolutionary rate comparisons with human orthologs revealed complementary trends in selective pressures. Genes slow-evolving in *S. microphthalmus* but fast-evolving in humans were enriched for core functions such as transcription and metabolism, indicating stronger purifying selection on essential cellular pathways in the energetically limited cave environment. Conversely, genes with accelerated divergence in *S. microphthalmus* yet slow in humans were enriched for transcriptional regulation, proteolysis and ion transport—functions also implicated in regions under selection (Section 4.2). Strikingly, homoeolog pairs with asymmetric evolutionary rates often exhibited one copy retaining a human-like rate while the other diverged, a pattern suggestive of early subfunctionalisation or neofunctionalisation. Although our data do not directly demonstrate gene loss or inactivation, these rate asymmetries point to distinct selective regimes that may precede functional divergence. Such divergent evolutionary fates for duplicate genes, enabled by allopolyploidy, likely supply a rich source of adaptive novelty (Birchler and Yang 2022) and constitute an important element of this species' diversification.

4.5 | A Diversification Scenario for an Allotetraploid Cavefish

Our collective findings allow us to propose a plausible scenario for the diversification of *S. microphthalmus*. In this view, allopatric isolation in fragmented cave habitats began within the last ~240,000 years, likely driven by Pleistocene climatic shifts. These isolated populations then followed distinct demographic trajectories, with some (e.g., the gracile morphotypes LLP and SCP) experiencing prolonged small N_e that accelerated divergence through genetic drift, while others (like the robust morphotype ZZP) persisted in more dynamic environments. Layered upon this demographic backdrop, the allotetraploid genome acted as an evolutionary catalyst. The functional redundancy of its duplicated subgenomes permitted differential evolutionary trajectories; we observe a general pattern where the H subgenome contributes more to broad-scale adaptive divergence, while the L copy retains focal adaptive diversity. This partitioning of roles, coupled with local adaptation to specific cave environments (evidenced by SOCs and selective sweeps), likely accelerated genomic and phenotypic differentiation among lineages, placing them on distinct evolutionary paths that represent the early stages of speciation.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | Conclusions

This study provides a comprehensive genomic perspective on the cave adaptation of *S. microphthalmus*, revealing how allopatric divergence, local selection and neutral processes interact to drive diversification. Our findings demonstrate substantial genetic differentiation among cave populations, shaped by both distinct demographic histories and population-specific selective pressures. While relaxed selection on visual genes emerges as a prominent feature, positive selection on other loci highlights a complex adaptive landscape. Crucially, our results suggest that the allotetraploid genome acts as an evolutionary catalyst. By enabling subgenome partitioning and diverse evolutionary fates for duplicated genes, allopolyploidy appears to accelerate adaptive divergence and provide multiple pathways toward speciation in this extreme environment. Future functional and transcriptomic studies will be essential to test the specific mechanisms outlined in our proposed diversification scenario and to fully understand the role of polyploidy in animal evolution.

Author Contributions

B.G. conceived and designed the research. Y.W., B.G., S.M.-M. and X.L. analysed the data. X.L., F.M., C.W. and Y.Z. contributed reagents, materials, or analytical tools. S.M.-M. wrote the paper under the leadership of X.L. and B.G. and with support from Y.W. All authors contributed to discussions and provided comments on the manuscript.

Acknowledgements

We are grateful to Haibo Liu, Yang Zhao, Jie Bai and Chengyi Niu for their assistance with sampling, and to Bohao Fang and Petri Kemppainen for help with data analyses. We also appreciate the feedback from the members of the Guo laboratory and Carles Vilà on early drafts of the manuscript. S.M.-M. extends special thanks to Xinxin Wu for the support provided throughout the writing of this work. This study was supported by grants from the Ministry of Science and Technology of the People's Republic of China (the National Key Program of Research and Development, 2023YFF1304800 and the Foreign Expert Project, QN2023061006L), the Institute of Zoology, Chinese Academy of Sciences (2023IOZ0104), State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management (SKLA2502 and an open funding grant), the National Natural Science Foundation of China (31970382, 32022009, W2412126 and 32150410358) and the Chinese Academy of Sciences (President's International Fellowship Initiative for Visiting Scientists, 2021PB0022).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The *Sinocyclocheilus tileihornes* reference genome is available in the Genome Warehouse (https://ngdc.cncb.ac.cn/gwh/) at the China National Center for Bioinformation (CNCB) under accession number GWHFQVD00000000.1. Mitochondrial protein-coding gene sequences for *Sinocyclocheilus guilinensis* and *S. tileihornes* outgroups are available in GenBase (https://ngdc.cncb.ac.cn/genbase/) under accessions ranging from C_AA105729.1 to C_AA105741.1 and from C_AA105742.1 to C_AA105754.1, respectively. Raw sequencing reads for the 47 *Sinocyclocheilus microphthalmus* study individuals and the 5 *S. tileihornes* outgroup samples have been deposited in the Genome Sequence Archive (GSA; https://ngdc.cncb.ac.cn/gsa/) under accession number CRA023833.

References

Alachiotis, N., and P. Pavlidis. 2018. "RAiSD Detects Positive Selection Based on Multiple Signatures of a Selective Sweep and SNP Vectors." *Communications Biology* 1: 79. https://doi.org/10.1038/s42003-018-0085-8.

Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19: 1655–1664. https://doi.org/10.1101/gr.094052.109.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Balogh, A., L. Ngo, K. S. Zigler, and G. Dixon. 2020. "Population Genomics in Two Cave-Obligate Invertebrates Confirms Extremely Limited Dispersal Between Caves." *Scientific Reports* 10: 17554. https://doi.org/10.1038/s41598-020-74508-9.

Bernt, M., A. Donath, F. Jühling, et al. 2013. "MITOS: Improved De Novo Metazoan Mitochondrial Genome Annotation." *Molecular Phylogenetics and Evolution* 69: 313–319. https://doi.org/10.1016/j.ympev.2012.08.023.

Birchler, J. A., and H. Yang. 2022. "The Multiple Fates of Gene Duplications: Deletion, Hypofunctionalization, Subfunctionalization, Neofunctionalization, Dosage Balance Constraints, and Neutral Variation." *Plant Cell* 34: 2466–2474. https://doi.org/10.1093/plcell/koac076.

Blanc, G., and K. H. Wolfe. 2004. "Functional Divergence of Duplicated Genes Formed by Polyploidy During Arabidopsis Evolution." *Plant Cell* 16: 1679–1691. https://doi.org/10.1105/tpc.021410.

Buhay, J. E., and K. A. Crandall. 2005. "Subterranean Phylogeography of Freshwater Crayfishes Shows Extensive Gene Flow and Surprisingly Large Population Sizes: Subterranean Phylogeography of Crayfishes." *Molecular Ecology* 14: 4259–4273. https://doi.org/10.1111/j.1365-294X.2005.02755.x.

Calderoni, L., O. Rota-Stabelli, E. Frigato, et al. 2016. "Relaxed Selective Constraints Drove Functional Modifications in Peripheral Photoreception of the Cavefish *P. andruzzii* and Provide Insight Into the Time of Cave Colonization." *Heredity* 117: 383–392. https://doi.org/10.1038/hdy.2016.59.

Cartwright, R. A., R. S. Schwartz, A. L. Merry, and M. M. Howell. 2017. "The Importance of Selection in the Evolution of Blindness in Cavefish." *BMC Evolutionary Biology* 17: 45. https://doi.org/10.1186/s12862-017-0876-4.

Charlesworth, D. 2006. "Balancing Selection and Its Effects on Sequences in Nearby Genome Regions." *PLoS Genetics* 2: e64. https://doi.org/10.1371/journal.pgen.0020064.

Cruickshank, T. E., and M. W. Hahn. 2014. "Reanalysis Suggests That Genomic Islands of Speciation Are due to Reduced Diversity, Not Reduced Gene Flow." *Molecular Ecology* 23: 3133–3157. https://doi.org/10.1111/mec.12796.

Culver, D. C., J. E. Kowalko, and T. Pipan. 2023. "Natural Selection Versus Neutral Mutation in the Evolution of Subterranean Life: A False Dichotomy?" *Frontiers in Ecology and Evolution* 11: 1080503.

Culver, D. C., and T. Pipan. 2019. *The Biology of Caves and Other Subterranean Habitats*. 2nd ed. Oxford University Press. https://doi.org/10.1093/oso/9780198820765.001.0001.

Cvekl, A., and W.-L. Wang. 2009. "Retinoic Acid Signaling in Mammalian Eye Development." *Experimental Eye Research* 89: 280–291. https://doi.org/10.1016/j.exer.2009.04.012.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10: giab008. https://doi.org/10.1093/gigascience/giab008.

Edger, P. P., and J. C. Pires. 2009. "Gene and Genome Duplications: The Impact of Dosage-Sensitivity on the Fate of Nuclear Genes."

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chromosome Research 17: 699–717. https://doi.org/10.1007/s10577-009-9055-9.

Fang, B., P. Kemppainen, P. Momigliano, X. Feng, and J. Merilä. 2020. "On the Causes of Geographically Heterogeneous Parallel Evolution in Sticklebacks." *Nature Ecology & Evolution* 4: 1105–1115. https://doi.org/10.1038/s41559-020-1222-6.

Frankham, R., J. D. Ballou, D. A. Briscoe, and K. H. McInnes. 2002. *Introduction to Conservation Genetics*. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511808999.

Fumey, J., H. Hinaux, C. Noirot, C. Thermes, S. Rétaux, and D. Casane. 2018. "Evidence for Late Pleistocene Origin of *Astyanax mexicanus* Cavefish." *BMC Evolutionary Biology* 18: 43. https://doi.org/10.1186/s12862-018-1156-7.

Gundemir, S., G. Colak, J. Tucholski, and G. V. W. Johnson. 2012. "Transglutaminase 2: A Molecular Swiss Army Knife." *Biochimica et Biophysica Acta* 1823: 406–419. https://doi.org/10.1016/j.bbamcr.2011.09.012.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35: 518–522. https://doi.org/10.1093/molbev/msx281.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. "Population Genomics of Parallel Adaptation in Threespine Stickleback Using Sequenced RAD Tags." *PLoS Genetics* 6: e1000862. https://doi.org/10.1371/journal.pgen.1000862.

IUCN. 1996. "*Sinocyclocheilus microphthalmus*. The IUCN Red List of Threatened Species." e.T20254A9181750. https://doi.org/10.2305/IUCN.UK.1996.RLTS.T20254A9181750.en.

Jakobsson, M., and N. A. Rosenberg. 2007. "CLUMPP: A Cluster Matching and Permutation Program for Dealing With Label Switching and Multimodality in Analysis of Population Structure." *Bioinformatics* 23: 1801–1806. https://doi.org/10.1093/bioinformatics/btm233.

Jeffery, W. R. 2001. "Cavefish as a Model System in Evolutionary Developmental Biology." *Developmental Biology* 231: 1–12. https://doi.org/10.1006/dbio.2000.0121.

Jeffery, W. R. 2009. "Regressive Evolution in *Astyanax* Cavefish." *Annual Review of Genetics* 43: 25–47. https://doi.org/10.1146/annurev-genet-102108-134216.

Jin, L., Z. Li, C. Wang, et al. 2022. "Contrasting Population Differentiation in Two Sympatric *Triplophysa* Loaches on the Qinghai–Tibet Plateau." *Frontiers in Genetics* 13: 958076. https://doi.org/10.3389/fgene.2022.958076.

Jones, F. C., M. G. Grabherr, Y. F. Chan, et al. 2012. "The Genomic Basis of Adaptive Evolution in Threespine Sticklebacks." *Nature* 484: 55–61. https://doi.org/10.1038/nature10944.

Juan, C., M. T. Guzik, D. Jaume, and S. J. B. Cooper. 2010. "Evolution in Caves: Darwin's 'Wrecks of Ancient Life' in the Molecular Era." *Molecular Ecology* 19: 3865–3880. https://doi.org/10.1111/j.1365-294X.2010.04759.x.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14: 587–589. https://doi.org/10.1038/nmeth.4285.

Kemppainen, P., C. G. Knight, D. K. Sarma, et al. 2015. "Linkage Disequilibrium Network Analysis (LDna) Gives a Global View of Chromosomal Inversions, Local Adaptation and Geographic Structure." *Molecular Ecology Resources* 15: 1031–1045. https://doi.org/10.1111/1755-0998.12369.

Kitahashi, T., D. Kurokawa, S. Ogiso, N. Suzuki, and H. Ando. 2021. "Light-Induced and Circadian Expressions of Melanopsin Genes opn4xa and opn4xb in the Eyes of Juvenile Grass Puffer *Takifugu alboplumbeus*." *Fish Physiology and Biochemistry* 47: 191–202. https://doi.org/10.1007/s10695-020-00901-w.

Kolberg, L., U. Raudvere, I. Kuzmin, J. Vilo, and H. Peterson. 2020. "gprofiler2 – An R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset g:Profiler." *F1000Research* 9: 709. https://doi.org/10.12688/f1000research.24956.2.

Lahti, D. C., N. A. Johnson, B. C. Ajie, et al. 2009. "Relaxed Selection in the Wild." *Trends in Ecology & Evolution* 24: 487–496. https://doi.org/10.1016/j.tree.2009.03.010.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., and R. Durbin. 2011. "Inference of Human Population History From Individual Whole-Genome Sequences." *Nature* 475: 493–496. https://doi.org/10.1038/nature10231.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, X., and B. Guo. 2020. "Substantially Adaptive Potential in Polyploid Cyprinid Fishes: Evidence From Biogeographic, Phylogenetic and Genomic Studies." *Proceedings of the Royal Society B: Biological Sciences* 287: 20193008.

Li, Z., B. Guo, J. Li, S. He, and Y. Chen. 2008. "Bayesian Mixed Models and Divergence Time Estimation of Chinese Cavefishes (Cyprinidae: *Sinocyclocheilus*)." *Scientific Bulletin* 53: 2342–2352. https://doi.org/10.1007/s11434-008-0297-2.

Lynch, M. 2007. *The Origins of Genome Architecture*. 1st ed. Sinauer Associates Inc.

Mack, K. L., J. B. Jaggard, J. L. Persons, et al. 2021. "Repeated Evolution of Circadian Clock Dysregulation in Cavefish Populations." *PLoS Genetics* 17: e1009642. https://doi.org/10.1371/journal.pgen.1009642.

Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov. 2021. "BUSCO Update: Novel and Streamlined Workflows Along With Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes." *Molecular Biology and Evolution* 38: 4647–4654. https://doi.org/10.1093/molbev/msab199.

Mao, T., Y. Liu, M. M. Vasconcellos, et al. 2025. "Caves as Species Pumps: Key Innovations, Isolation and Periodic Introgression Drive the World's Largest Cavefish Radiation in a Dynamic Karstic Landscape." *Molecular Ecology* 34: e70038. https://doi.org/10.1111/mec.70038.

Marçais, G., A. L. Delcher, A. M. Phillippy, R. Coston, S. L. Salzberg, and A. Zimin. 2018. "MUMmer4: A Fast and Versatile Genome Alignment System." *PLoS Computational Biology* 14: e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Mayr, E. 1942. *Systematics and the Origin of Species From the Viewpoint of a Zoologist*. Columbia University Press.

McKenna, A., M. Hanna, E. Banks, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data." *Genome Research* 20: 1297–1303. https://doi.org/10.1101/gr.107524.110.

Meng, G., Y. Li, C. Yang, and S. Liu. 2019. "MitoZ: A Toolkit for Animal Mitochondrial Genome Assembly, Annotation and Visualization." *Nucleic Acids Research* 47: e63. https://doi.org/10.1093/nar/gkz173.

Minh, B. Q., H. A. Schmidt, O. Chernomor, et al. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Montero-Mendieta, S., K. Tan, M. J. Christmas, et al. 2019. "The Genomic Basis of Adaptation to High-Altitude Habitats in the Eastern Honey Bee (*Apis cerana*)." *Molecular Ecology* 28: 746–760. https://doi.org/10.1111/mec.14986.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Moran, D., R. Softley, and E. J. Warrant. 2015. "The Energetic Cost of Vision and the Evolution of Eyeless Mexican Cavefish." *Science Advances* 1: e1500363. https://doi.org/10.1126/sciadv.1500363.

Nei, M. 2005. "Bottlenecks, Genetic Polymorphism and Speciation." *Genetics* 170: 1–4. https://doi.org/10.1093/genetics/170.1.1.

Noor, M. A. F., and S. M. Bennett. 2009. "Islands of Speciation or Mirages in the Desert? Examining the Role of Restricted Recombination in Maintaining Species." *Heredity* 103: 439–444. https://doi.org/10.1038/hdy.2009.151.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. "Divergent Selection and Heterogeneous Genomic Divergence." *Molecular Ecology* 18: 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x.

Orrenius, S., B. Zhivotovsky, and P. Nicotera. 2003. "Regulation of Cell Death: The Calcium–Apoptosis Link." *Nature Reviews. Molecular Cell Biology* 4: 552–565. https://doi.org/10.1038/nrm1150.

Otto, S. P., and J. Whitton. 2000. "Polyploid Incidence and Evolution." *Annual Review of Genetics* 34: 401–437. https://doi.org/10.1146/annurev.genet.34.1.401.

Pavlidis, P., D. Živković, A. Stamatakis, and N. Alachiotis. 2013. "SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes." *Molecular Biology and Evolution* 30: 2224–2234. https://doi.org/10.1093/molbev/mst112.

Peuß, R., A. C. Box, S. Chen, et al. 2020. "Adaptation to Low Parasite Abundance Affects Immune Investment and Immunopathological Responses of Cavefish." *Nature Ecology & Evolution* 4: 1416–1430. https://doi.org/10.1038/s41559-020-1234-2.

Protas, M. E., C. Hersey, D. Kochanek, et al. 2006. "Genetic Analysis of Cavefish Reveals Molecular Convergence in the Evolution of Albinism." *Nature Genetics* 38: 107–111. https://doi.org/10.1038/ng1700.

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81: 559–575. https://doi.org/10.1086/519795.

Ronfort, J. 1999. "The Mutation Load Under Tetrasomic Inheritance and Its Consequences for the Evolution of the Selfing Rate in Autotetraploid Species." *Genetical Research* 74: 31–42. https://doi.org/10.1017/S0016672399003845.

Sajovic, J., A. Meglič, D. Glavač, Š. Markelj, M. Hawlina, and A. Fakin. 2022. "The Role of Vitamin A in Retinal Diseases." *International Journal of Molecular Sciences* 23: 1014. https://doi.org/10.3390/ijms23031014.

Sbordoni, V., G. Allegrucci, and D. Cesaroni. 2012. "Population Structure." In *Encyclopedia of Caves*, 608–618. Elsevier/Academic Press.

Slatkin, M. 2008. "Linkage Disequilibrium— Understanding the Evolutionary Past and Mapping the Medical Future." *Nature Reviews. Genetics* 9: 477–485. https://doi.org/10.1038/nrg2361.

Smith, M. L., and M. W. Hahn. 2021. "New Approaches for Inferring Phylogenies in the Presence of Paralogs." *Trends in Genetics* 37: 174–187. https://doi.org/10.1016/j.tig.2020.08.012.

Soltis, D. E., and P. S. Soltis. 1999. "Polyploidy: Recurrent Formation and Genome Evolution." *Trends in Ecology & Evolution* 14: 348–352. https://doi.org/10.1016/S0169-5347(99)01638-9.

Soltis, P. S., and D. E. Soltis. 2012. *Polyploidy and Genome Evolution*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31442-1.

The Chimpanzee Sequencing and Analysis Consortium. 2005. "Initial Sequence of the Chimpanzee Genome and Comparison With the Human Genome." *Nature* 437: 69–87. https://doi.org/10.1038/nature04072.

Trontelj, P. 2019. "Adaptation and Natural Selection in Caves." In *Encyclopedia of Caves*, 40–46. Elsevier. https://doi.org/10.1016/B978-0-12-814124-3.00006-6.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. "Genomic Islands of Speciation in *Anopheles gambiae*." *PLoS Biology* 3: e285. https://doi.org/10.1371/journal.pbio.0030285.

Van De Peer, Y., E. Mizrachi, and K. Marchal. 2017. "The Evolutionary Significance of Polyploidy." *Nature Reviews. Genetics* 18: 411–424. https://doi.org/10.1038/nrg.2017.26.

Wang, D., Y. Zhang, Z. Zhang, J. Zhu, and J. Yu. 2010. "KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies." *Genomics, Proteomics & Bioinformatics* 8: 77–80. https://doi.org/10.1016/S1672-0229(10)60008-3.

Wang, K., M. Li, and H. Hakonarson. 2010. "ANNOVAR: Functional Annotation of Genetic Variants From High-Throughput Sequencing Data." *Nucleic Acids Research* 38: e164. https://doi.org/10.1093/nar/gkq603.

Wang, M., X. Li, C. Wang, et al. 2024. "Asymmetric and Parallel Subgenome Selection Co-Shape Common Carp Domestication." *BMC Biology* 22: 4. https://doi.org/10.1186/s12915-023-01806-9.

Wang, Y., H. Tang, J. D. Debarry, et al. 2012. "MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity." *Nucleic Acids Research* 40: e49. https://doi.org/10.1093/nar/gkr1293.

Wilkens, H. 1988. "Evolution and Genetics of Epigean and Cave *Astyanax fasciatus* (Characidae, Pisces)." In *Evolutionary Biology: Volume 23*, edited by M. K. Hecht and B. Wallace, 271–367. Springer US. https://doi.org/10.1007/978-1-4613-1043-3_8.

Wolf, J. B. W., and H. Ellegren. 2017. "Making Sense of Genomic Islands of Differentiation in Light of Speciation." *Nature Reviews. Genetics* 18: 87–100. https://doi.org/10.1038/nrg.2016.133.

Wolfe, K. H. 2001. "Yesterday's Polyploids and the Mystery of Diploidization." *Nature Reviews. Genetics* 2: 333–341. https://doi.org/10.1038/35072009.

Yang, J., X. Chen, J. Bai, et al. 2016. "The *Sinocyclocheilus* Cavefish Genome Provides Insights Into Cave Adaptation." *BMC Biology* 14: 1. https://doi.org/10.1186/s12915-015-0223-4.

Yang, L., T. Sado, M. Vincent Hirt, et al. 2015. "Phylogeny and Polyploidy: Resolving the Classification of Cyprinine Fishes (Teleostei: Cypriniformes)." *Molecular Phylogenetics and Evolution* 85: 97–116. https://doi.org/10.1016/j.ympev.2015.01.014.

Yang, Z. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24: 1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang, Z., and J. P. Bielawski. 2000. "Statistical Methods for Detecting Molecular Adaptation." *Trends in Ecology & Evolution* 15: 496–503. https://doi.org/10.1016/S0169-5347(00)01994-7.

Yoshizawa, M., B. G. Robinson, E. R. Duboué, et al. 2015. "Distinct Genetic Architecture Underlies the Emergence of Sleep Loss and Prey-Seeking Behavior in the Mexican Cavefish." *BMC Biology* 13: 15. https://doi.org/10.1186/s12915-015-0119-3.

Zhang, C., S.-S. Dong, J.-Y. Xu, W.-M. He, and T.-L. Yang. 2019. "PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files." *Bioinformatics* 35: 1786–1788. https://doi.org/10.1093/bioinformatics/bty875.

Zhang, M., Q. Zhao, J. Huang, et al. 2025. "A Splicing Mutation in *Mitfa* Is Involved in the Depigmentation of Cavefish *Triplophysa rosa*." *Molecular Biology and Evolution* 42: msaf175. https://doi.org/10.1093/molbev/msaf175.

Zhao, Q., F. Shao, Y. Li, S. V. Yi, and Z. Peng. 2022. "Novel Genome Sequence of Chinese Cavefish (*Triplophysa rosa*) Reveals Pervasive

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Relaxation of Natural Selection in Cavefish Genomes." *Molecular Ecology* 31: 5831–5845. https://doi.org/10.1111/mec.16700.

Zhao, Y. 2021. "The Population Diversity and Conservation of Endangered Chinese Cavefish, *Sinocyclocheilus microphthalmus*." Master's thesis, Hebei University, Hebei, China.

Zhao, Y., C. Zhang, and G. Proudlove. 2021. *Fishes of the Genus Sinocyclocheilus (Cypriniformes: Cyprinidae) in China: Systematics, Biology, Biogeography and Cave Adaptation*. Upfront Publishing.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figures S1–S20:** mec70118-sup-0001-FigureS1-S20.pdf. **Tables S1–S13:** mec70118-sup-0002-TableS1-S13.xlsx.

1365294x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/mec.70118 by <Shibboleth>-member@ioz.ac.cn, Wiley Online Library on [25/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License