Check for updates

EVOLUTIONARY BIOLOGY

Insights into the genetic basis of bilateral head asymmetry in a scale-eating cichlid fish

Xiaomeng Tian[1], Ming Li[1], Axel Meyer[1,2]*

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Almost all animals are bilaterally symmetrical, and only rarely is this universality broken and maintained as a stable polymorphism. One such case is *Perissodus microlepis*, a scale-eating cichlid fish endemic to Lake Tanganyika. This species is polymorphic with left- or right-headed morphs, each specialized for scraping scales from the opposite side of their prey. This polymorphism is stably maintained around a 50:50 ratio and is likely driven by negative frequency–dependent selection, a classic form of balancing selection. Using a genome-wide dataset, including a de novo haplotype-resolved reference genome, whole-genome sequences, and microcomputed tomography scans of 102 individuals, we identified 72 genetic loci linked to head asymmetry. Nearby genes are enriched for roles in symmetry establishment, facial morphology, and nervous system function, with evidence suggesting that balancing selection maintains this diversity. Our study expands our understanding of the genetic basis of head asymmetry and provides candidate genes for further research.

INTRODUCTION

Ever since Darwin, evolutionary biologists have tried to understand the forces that shape the "beautiful and harmonious diversity of nature" (*1*). Of particular interest, because they are rare and hard to explain, are cases of distinct intraspecific phenotypes called polymorphisms (*2*). The origin and maintenance of polymorphisms remain a central, unsolved question in evolutionary biology (*3*–*8*), because they exist despite theory predicting that directional natural selection and genetic drift might be expected to erode such diversity (*9*, *10*). Balancing selection would favor the continued existence of both phenotypes and has been proposed as one potential mechanism for how genetic variation is maintained in populations (*11*–*13*). *Perissodus microlepis*, an endemic cichlid fish from Lake Tanganyika, is considered a textbook example for balancing selection (*14*, *15*). This species is characterized by extreme trophic adaptations—lateralized head asymmetry that facilitates feeding on the scales on one or the other side of their prey fish (*16*–*21*). The head asymmetry is coupled with lateralized foraging behavior, where left-headed morphs (L-morphs) preferentially feed on the scales of the right side of prey, and vice versa for right-headed morphs (R-morphs) (*22*, *23*). The relative frequency of the two morphs in natural populations cycles every 4 to 5 years but remains stable in the long term at ~50:50 (*14*, *16*, *17*). This is apparently due to the higher fitness of the rare morph relative to the abundant morph, because prey will pay more attention toward the side that is more frequently attacked, diminishing the foraging success of the more abundant morph, thereby reducing its predation success and ultimately its fitness (*24*, *25*).

The genetic basis of morphological asymmetry in general and in *P. microlepis* in particular has been a subject of interest and debate for nearly three decades. This phenotype was initially thought to be a simple Mendelian trait with bimodal distribution (*14*, *26*). However, subsequent studies found a more unimodal distribution with a preponderance of symmetrical heads in some populations (*17*, *27*) and failed to validate the originally proposed Mendelian locus model (*16*). Subsequently, a genome-wide study using double-digest restriction site–associated DNA (ddRAD) sequencing identified a single single-nucleotide polymorphism (SNP) that was significantly differentiated between L- and R-morphs (*28*) but only exhibited a weak effect on asymmetry, providing further evidence for a quantitative basis of asymmetry (*29*). However, the limited resolution of ddRAD sequencing (i.e., <5% of the genome sampled) and the lack of a reference genome for this species prevented further characterizations of the genetic basis of this phenotype. Adding to the uncertainty is the question of to what extent head asymmetry is genetically determined (*22*). A previous study proposed that morphological asymmetry in *P. microlepis* is likely enhanced by developmental plasticity via lateralized feeding behavior contributing to overall trait variation (*19*). Heritability estimates suggest that the variation in head asymmetry has some degree of genetic contribution, with a field estimate of 0.22 (*30*). Experimental investigations into the development of lateralized predation behavior demonstrate that the behavioral laterality of this fish is a combination of genetics and behavioral experiences (*31*). In addition, research on the brain transcriptome suggests that handedness in predation behavior may partially result from lateralized gene expressions in the brain (*32*). Together, these findings indicate that both morphological and behavioral asymmetries have a measurable genetic contribution and might share a common or interconnected genetic basis. Advances in morphological quantification methods and genomic technologies now offer promising opportunities for further elucidating the genetic mechanisms underlying this complex trait.

A preliminary study based on microcomputed tomography (μCT) revealed that some of the craniofacial structures exhibited a consistent bimodal pattern that differed between the L- and R-morphs (*21*). On the basis of this, we proposed three hypotheses: (i) Head asymmetry is determined by a few major effect loci; (ii) behavioral laterality (determined by a few major effect loci) shapes head asymmetry through developmental plasticity; (iii) head asymmetry and behavioral laterality interact and are determined by a large number of small effect loci. To test these hypotheses, we first generated the first chromosome-level reference genome for this species. In addition, we resequenced the genomes of 102 wild-caught individuals and identified the genomic signatures of balancing selection through various population genetic methods. Furthermore, we conducted a

[1]Department of Biology, University of Konstanz, Konstanz, Germany. [2]Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA.
*Corresponding author. Email: axel.meyer@uni-konstanz.de

Downloaded from https://www.science.org on July 31, 2025

genome-wide association study (GWAS) to test hypotheses about the genetic basis of head asymmetry. Our study investigated the long-standing puzzle concerning the genetic basis of bilateral head asymmetry. We also highlight the role of balancing selection in maintaining head asymmetry polymorphism as part of ecological adaptation.

RESULTS

Chromosome-level assembly of the *P. microlepis* genome

To better understand the genetic basis of bilateral head asymmetry in *P. microlepis*, we created a high-quality chromosome-level reference genome for this species. Specifically, we used PacBio HiFi long reads and chromosome conformation capture (Hi-C) Illumina read pairs to produce two high-quality haplotype-resolved draft genomes with N50 contig lengths of 15.13 and 10.83 Mb, respectively. In addition, we further used Hi-C Illumina read pairs to scaffold the contigs of the two haplotype-resolved draft genomes. The resulting two haplotypes are of comparable quality (table S1), with the more complete one spanning 853.31 Mb with a GC content of 40.71% and a scaffold N50 of 37.72 Mb and serving as the reference genome for subsequent analyses (Fig. 1A). Our Hi-C analyses scaffolded the *P. microlepis* genome, anchoring 99.33% of the genome assembly into 22 pseudochromosomes (Fig. 1B). Comparative analysis with newly published other African cichlid genomes indicates that our assembly exhibits higher contig N50 values, surpassing the published *P. microlepis* genome (contig N50: 9.20 kb; GCA_015105195.1) assembled solely from short-read data by three orders of magnitude (Fig. 1C). Furthermore, the *Perissodus* genome is nearly 100% complete, as evidenced by the BUSCO score of 99.56%, and is among the best in genome completeness in cichlids (Fig. 1D).

On the basis of transcriptome evidence and homologous proteins of vertebrates, we identified 26,393 protein-coding genes in the *P. microlepis* genome. This is comparable to the number of genes in other African cichlid fish genomes that have been annotated at National Center for Biotechnology Information [NCBI; e.g., the NCBI Annotation Release 104 of *Maylandia zebra* (GCF_000238955.4): 25,898 protein-coding genes; the NCBI Annotation Release 104 of *Oreochromis niloticus* (GCF_001858045.2): 29,550 protein-coding

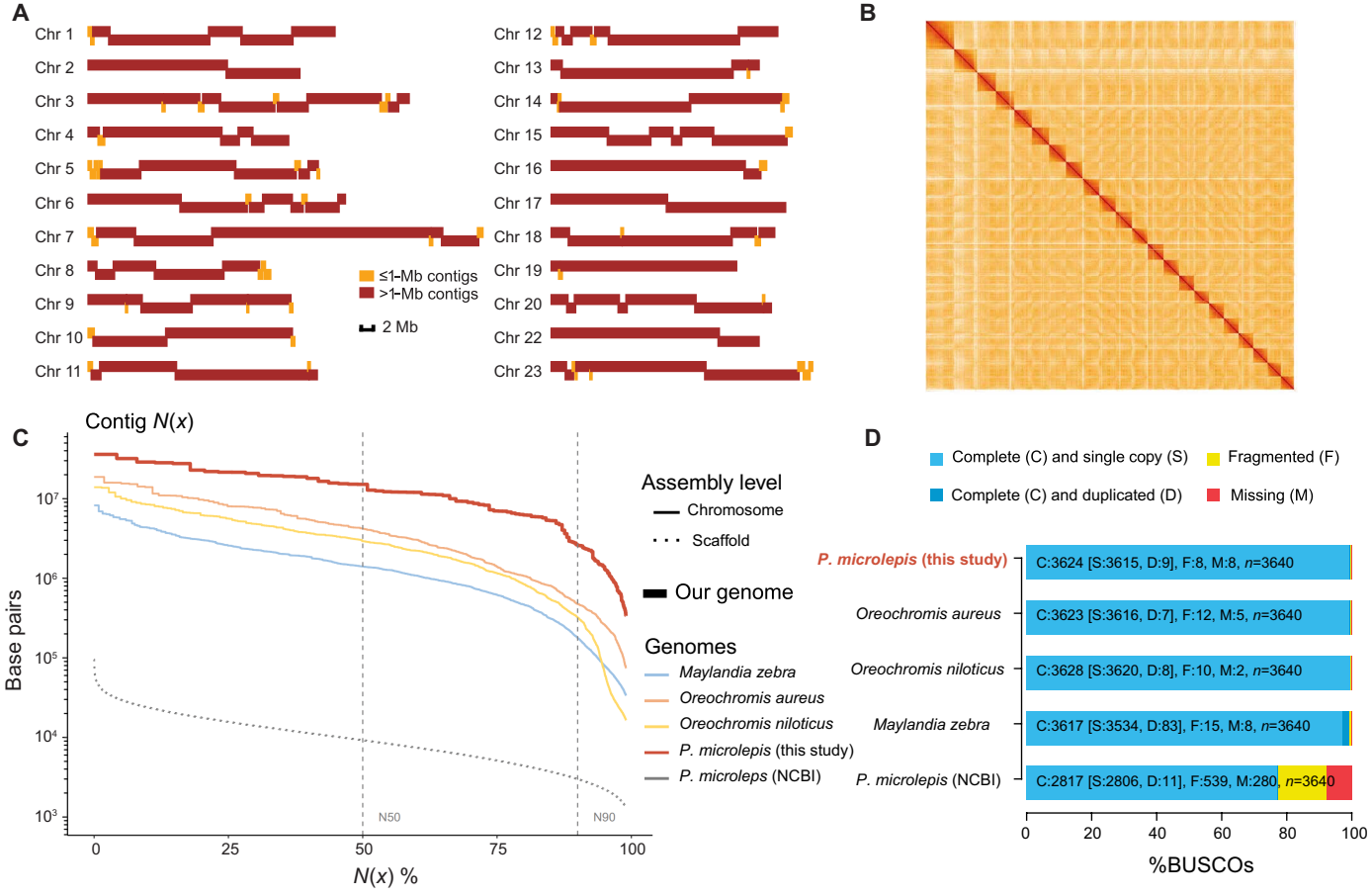

Fig. 1. De novo chromosome-level reference genome for *P. microlepis*. (**A**) Overview of contigs in the *P. microlepis* reference genome. Each bar represents a continuous gap-free contig, with orange representing contigs shorter than 1 Mb and red representing contigs longer than 1 Mb. (**B**) Hi-C contact map with chromosomes displayed in size order from top to bottom and from left to right. (**C**) Comparison of assembly contiguity. *N*(*x*) % graphs show contig sizes (*y* axis), in which the *x* percent of the assembly consists of contigs of at least that size. The gray dotted line represents the previous *P. microlepis* reference genome at the scaffold level (GenBank: GCA_015105195.1). The thin colorful lines represent the available chromosome-level reference genomes of other African cichlids, while the bold red line shows the newly generated chromosome-level reference genome for *P. microlepis* by this study. (**D**) Comparison of completeness of genome assembly as a percentage of 3640 actinopterygian_odb10 genes from BUSCO.

Downloaded from https://www.science.org on July 31, 2025

genes]. Meanwhile, 7904 genes are annotated with at least two alternatively spliced transcripts. BUSCO analysis suggested that 98.08% of the 3640 orthologs in the Actinopterygii dataset (actinopterygian_odb10) are completely annotated, reflecting a high-quality gene annotation. In addition, we predicted 305.84 Mb of repetitive sequences, accounting for 35.84% of the assembled genome, in which DNA transposons were the majority repeats (11.65%), followed by long interspersed nuclear elements (7.36%) and long terminal repeat elements (5.38%) (figs. S1 and S2).

Detecting the genomic signature of balancing selection

To test whether balancing selection contributed to the evolution and maintenance of key traits in this species, we assessed the genomic signatures of balancing selection by generating whole-genome data from 102 wild individuals of *P. microlepis* (table S2). The whole-genome sequenced samples have an average effective genome coverage of 27.6 ± 1.1× per individual, resulting in 2,727,283 high-quality SNPs after alignment to our new reference genome and necessary filtration (see Materials and Methods). To explicitly test for evidence of balancing selection, we applied two well-established methods—BetaScan (β) (*33*) and BalLeRMiX (B0,MAF) (*34*), each grounded in different principles on the resulting genome-wide variants (see Materials and Methods). For both methods, we deemed regions with scores surpassing the genome-wide 99th and 99.9th quantiles as significant. Using stringent criteria on β values, we detected 449 peaks (99th quantile) and 40 peaks (99.9th quantile) across the genome (Fig. 2A). The most substantial peak, spanning 5.3 kb and located on chromosome 3, harbored 221 SNPs with the highest β values. The identical criteria were used for B0,MAF values, resulting in the identification of 798 and 59 peaks, respectively. The longest peak is the same as the one found with the other pipeline, spanning 6.5 kb, which contained 284 SNPs with the highest B0,MAF values. A total of 367 peaks (99th quantile) and 27 peaks (99.9th quantile) was shared between these two methods, constituting 81.73 and 67.50% of total β peaks, respectively (data S1 and S2 and tables S3 and S4).

Hereafter, we focused on regions identified as exhibiting balancing selection signals by both methods. First, we examined their distribution between gene and nongene regions. Relative to the genome-wide average gene proportion (46.47%; including exons and introns), we observed a significantly lower occurrence of these

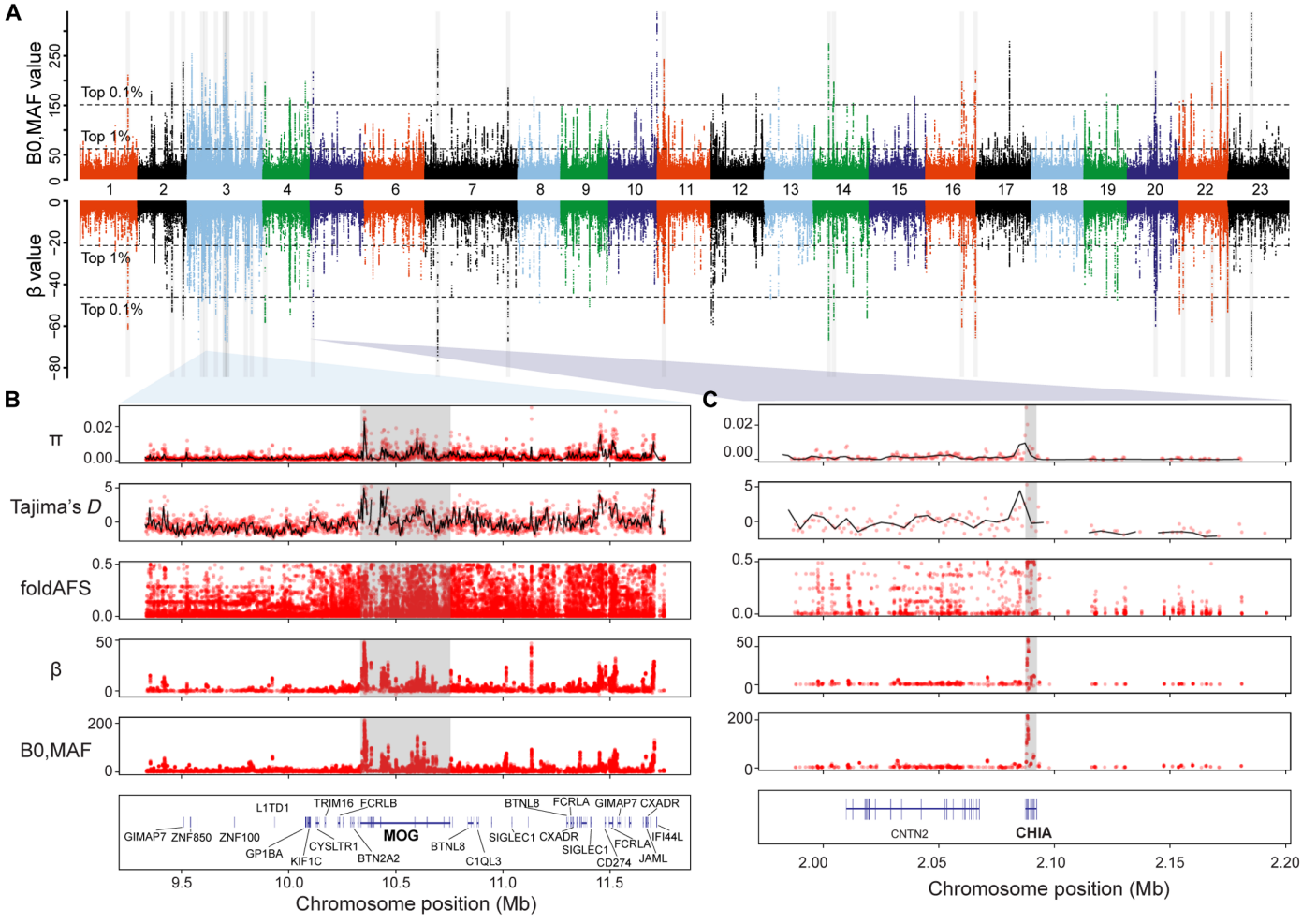

Fig. 2. Signatures of balancing selection in the *P. microlepis* genome. (**A**) Genome-wide signals of balancing selection revealed by B0,MAF (top panel) and β (bottom panel) scores. Quantiles of 1% and 1‰ are shown in the plot. (**B**) Genetic characteristics of a case gene—*MOG*. (**C**) Genetic characteristics of a case gene—*CHIA*. From top to bottom, the characteristics are Π, Tajima's *D*, foldAFS, β, and B0,MAF.

Downloaded from https://www.science.org on July 31, 2025

regions within gene regions (35.50%) (chi-square test, *P* value <0.05). Subsequently, we identified 113 genes (at the 99th percentile) and 9 genes (at the 99.9th percentile) that were fully or partially subjected to balancing selection. Among these, the nine genes were *SLC44A4*, *MOG*, *FCRLB*, *HIST2H2BE*, *HERC3*, *CHIA*, *PLCH2*, and two uncharacterized genes. Expectedly, the majority of these genes are involved in immune response, as immune genes are frequently reported to be subject to balancing selection because of pathogen polymorphism and other factors. Besides, *MOG* (Fig. 2B) is linked to the nervous system, which insulates nerve fibers in the central nervous system (*35*), and *CHIA* (Fig. 2C) is involved in chitin metabolism, a key component of fish scales (*36*, *37*). Following this, we conducted functional enrichment analysis on these genes, revealing enrichment in 28 Gene Ontology (GO) pathways. Among these, 19 were attributed to biological processes, 7 to molecular functions, and 1 to cellular components (fig. S3). Predominantly, these pathways are associated with immune response (e.g., GO:0001817, GO:0002376, GO:0019731, and GO:0001618), signal transmission (e.g., GO:0015220, GO:0004435, GO:0001817, and GO:0007417), cell growth (e.g., GO:0030307 and GO:0019900), and energy metabolism (GO:0044245) (fig. S3).

Additional statistics further reinforce the role of balancing selection in the two gene regions besides immune genes *MOG* and *CHIA*, because the immune system is not the focus of this study. We first dissected patterns of genetic diversity in genomic regions housing *MOG* and *CHIA* (Fig. 2, B and C), and we found that these two regions exhibit local Π peaks, which exceed chromosome-specific and genome-wide mean Π values by, on average, ~20-fold (Welch's *t* test, *P* value <0.05), which is in line with their folded allele frequency spectrum (foldAFS) where we saw clusters of intermediate frequencies of SNPs (Fig. 2, B and C). Furthermore, we found elevated Tajima's *D* peaks in these two genes, indicating excess intermediate-frequency alleles and the scarcity of rare alleles (Fig. 2, B and C). The Tajima's *D* peaks of *MOG* and *CHIA* stand out against negatively skewed genome-wide values and are significantly higher than chromosome background distribution by approximately sixfold on average (Welch's *t* test, *P* value <0.05). Collectively, candidate genes, such as *MOG* and *CHIA*, have among the highest genetic diversity (Π), more moderate allele frequency (foldAFS), and lowest segregating sites in the genome (Tajima's *D*), indicating a concentration of genetic diversity and the ongoing strong force of balancing selection in this species.

Head asymmetric quantification based on μCT

The μCT-based phenotypic quantification strategy was applied to 102 fish with genomic data, including the μCT data of 40 fish that we published earlier (*21*). For each fish head, we generated more than 1000 CT scans and isolated the specific ossified units of interest that are part of the functional suspensorium and kinetic jaw protrusion apparatus (fig. S4). Most samples showed a consistent deflection direction among the three craniofacial bones involved in the scale-eating process (premaxilla, maxilla, and mandible), suggesting a shared underlying mechanism (fig. S5). By more than doubling the sample size from 40 to 102, we found individuals with even higher levels of bone volumetric asymmetry (maximum 25% bigger on the expanded side than the smaller side). Regarding the distributions of volumetric asymmetry among the sampled individuals, we found a clearer pattern of bimodal distribution (Fig. 3, A to C, and fig. S6).

More detailed asymmetries were analyzed using a three-dimensional (3D) geometric morphometrics approach on 102 fish, with 96 landmarks placed on the head, including the premaxilla, maxilla, mandible, and suspensorium (fig. S4). The first three principal components in the 3D shape analyses accounted for 59.86%, 8.25% (20.55% after correction), and 6.15% (15.3% after correction) of the total variance (see Materials and Methods). PC3 primarily corresponded to head asymmetry and is significantly correlated with both field quantification and volume asymmetries (Fig. 3D; Pearson correlation, *P* values <0.05). We further quantified 313 distance asymmetries by connecting two landmarks on both sides of the same bone (Fig. 4, A to E). In terms of the trait distribution, 46 of them fit the bimodal distribution (kernel density estimation) and are strongly correlated with the volume asymmetries (Pearson correlation, *P* values <0.05). Among them, two are located on the premaxilla, one on the maxilla, nine on the mandible, and the rest on the suspensorium (figs. S7 to S12 and data S7). Nearly 80% of the 313 distance asymmetries are significantly correlated with each volume asymmetry (77.6% for the premaxilla, 80.2% for the maxilla, and 75.7% for the mandible) (data S3 to S5), indicating that the changes in these regions together contribute to the final morphological changes and possibly share a genetic and developmental basis for asymmetry.

GWAS analysis of head asymmetry

To study the genetic basis of bilateral head asymmetry, we performed a GWAS analysis on 316 morphological traits capturing the head asymmetry (313 distance asymmetries and 3 volumetric asymmetries). We identified 170 genome-wide significant variants [*P* value $<5 \times 10^{-8}$ and minor allele frequency (MAF) >0.05] in 72 loci for at least one morphological trait (Fig. 4F). Of these, 72 SNPs also passed the study-wide significance (*P* value $<1.58 \times 10^{-10}$). The SNPs passing the genome-wide threshold clumped on 34 linear distance asymmetries, located in the premaxilla, mandible, and suspensorium (data S6). Among them, 61 SNPs (35.88%) are significantly associated with more than one distance asymmetry on the same bone, indicating a shared genetic and developmental basis. Of these 170 SNPs, 57 are located in genic regions (including exons and introns), accounting for 33.53%, a percentage that is significantly lower than the ratio of gene region (46.37%) in the whole genome (chi-square test, *P* value <0.05). Only one SNP is located in an exon (synonymous variant on *ADCY5*), suggesting that the majority of these SNPs work in a regulatory manner (chi-square test, *P* value <0.05). Functional enrichment was performed on the 78 genes surrounding these SNPs. We found that the genes associated with the leading variants were highly enriched in biological processes of signaling, immune response, and regulation of cellular activities (table S5). We then looked in detail at the specific functions of each of these genes (table S6) and found two genes that directly play a role in left/right axis specification. *Notch1a* (one gene in the *Notch* signaling pathway) plays a primary role in establishing left-right asymmetry in mice by directly regulating *Nodal* gene expression (*38*). The lateralization of the vertebrate brain could be reversed by inactivation of *Notch* signaling in zebrafish (*39*). Moreover, the cadherin gene (*CDH7* in humans) can establish tissue-wide polarity by migrating cells in *Caenorhabditis elegans* (*40*). Evidence from both *Drosophila* and the chick highlights the function of cadherins in generating asymmetric tissue organization by setting up anterior-posterior and left-right asymmetry, respectively (*41*, *42*). In addition, we found that many of the GWAS-identified genes are involved in facial morphology (table S6). For instance, *LSP1* is also found in the large-scale GWAS of the human face morphology (*43*), and it has high expression

Downloaded from https://www.science.org on July 31, 2025

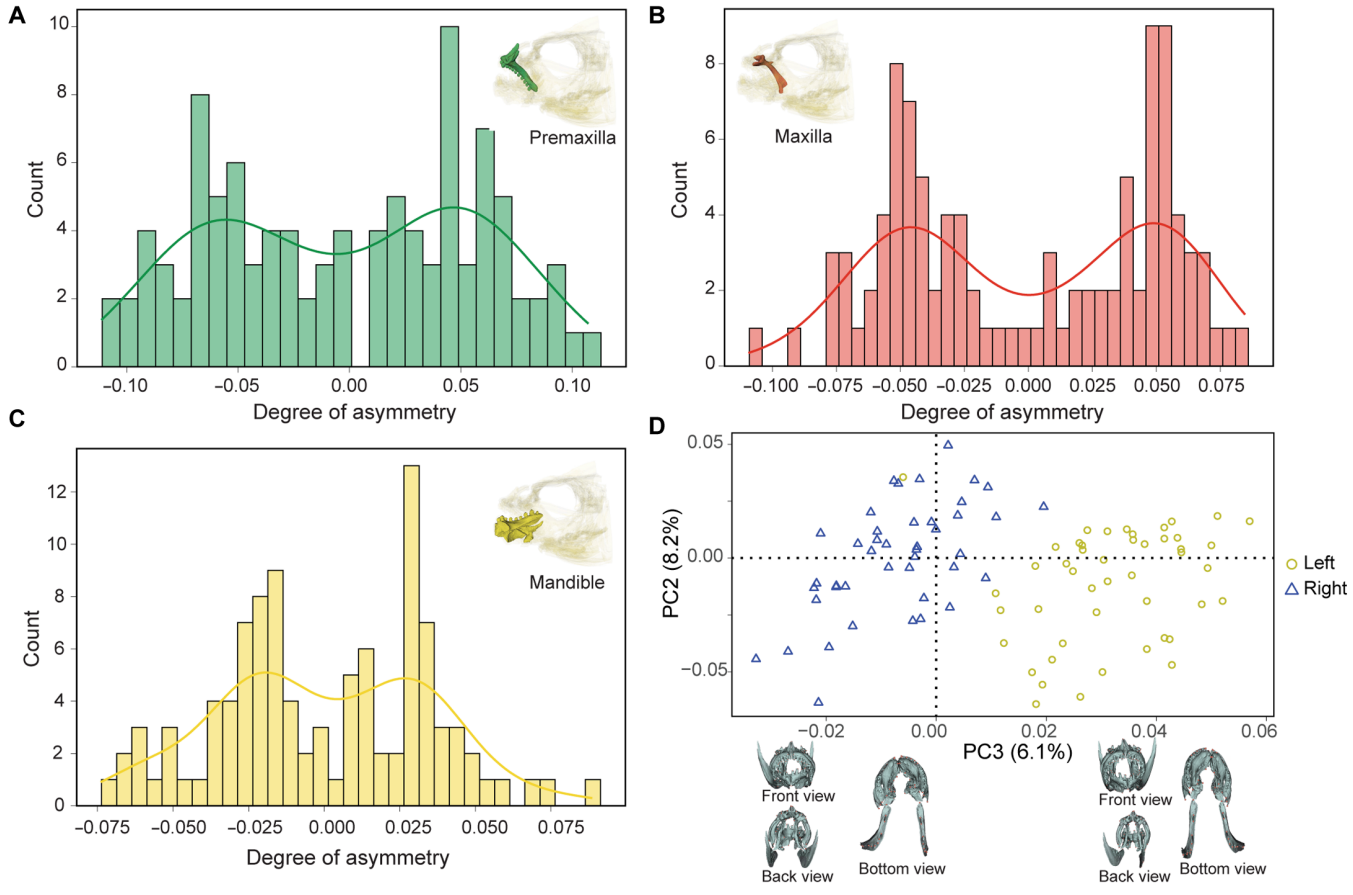

Fig. 3. Volume asymmetries of the craniofacial skeletons in 102 *P. microlepis* specimens. Distribution of the relative volumetric asymmetry between left and right sides of the premaxilla, maxilla, and mandible is displayed in (**A**) to (**C**), respectively. The relative differences between left and right sides were calculated on the basis of the equation Degree of asymmetry = (Left − Right)/(Left + Right) (see Materials and Methods). (**D**) Principal components analysis: PC2 and PC3 axes plotted. The classification of L- and R-morphs for the samples was determined on the basis of independent volume disparities comparing left and right sides.

in the bone marrow and skin (*44*). Silencing of *LSP1* reduced the migratory behavior of endothelial cells (*45*). *PLXNB1* has been reported to modulate osteogenesis by inhibiting the osteogenesis of osteoblasts (*46*). *PLXNB1*$^{-/-}$ mice had a high bone mass because of an increase in osteoblastic bone formation (*47*), and its overexpression promoted *BMSC* chondrogenic differentiation (*48*). *ERBB3* is an important member of the epidermal growth factor receptor family and may participate in tooth morphogenesis (*49*). Moreover, we found that a set of genes is involved in the nervous system, in particular *SLC6A13*, *TRRAP*, *TRIM67*, *CHD5*, *GRIP2*, *C9orf72*, and *NPTX2*, and is highly expressed in the brain, and their mutation or deletion can lead to neurological diseases (*50*–*56*). For instance, *TRRAP* is involved in neurodevelopmental disorders characterized by distinctive craniofacial features and mental retardation (*50*). *TRIM67* exhibits differential expression between the left and right telencephalon in chickens at 4 days posthatching (*57*). Members of the tripartite motif (*TRIM*) family have been linked to human handedness (e.g., *TRIM36*) (*58*) and the structural left-right asymmetry of the human brain (e.g., *TRIM14*) (*59*).

The GWAS signals are subject to strong balancing selection

To explore the hypothesis that balancing selection is contributing to the maintenance of the head asymmetry polymorphism in *P. microlepis*, we further accessed the characteristics of balancing selection on these GWAS signals. We identified, compared to the average of the whole genome, those 1-kb windows where these SNPs reside that have significantly higher nucleotide diversity (Π) (Fig. 5A). The Tajima's *D* values within the windows containing these SNPs are predominantly positive, indicating a pattern consistent with balancing selection (Fig. 5B). The frequencies of these SNPs tend to be intermediate, with a notable enrichment around 0.5 (Fig. 5C). In addition, two specific metrics for identifying balancing selection, β and B0,MAF, show elevated values in the GWAS signals (Fig. 5, D and E). Overall, from various analytical perspectives, we provide evidence that the signals identified by our GWAS exhibit stronger characteristics of balancing selection than the genome-wide background.

DISCUSSION

Our study dissected the genetic basis of breaking the bilateral symmetry in the textbook example *P. microlepis*. The frequencies of the two head morphs fluctuate, but because they always return to a 50:50 ratio, it is assumed that the maintenance of this polymorphism is due to negative frequency–dependent selection (*14*, *60*). By constructing the first chromosome-level reference genome and sequencing hundreds of individuals, we examined the genome-wide

Downloaded from https://www.science.org on July 31, 2025

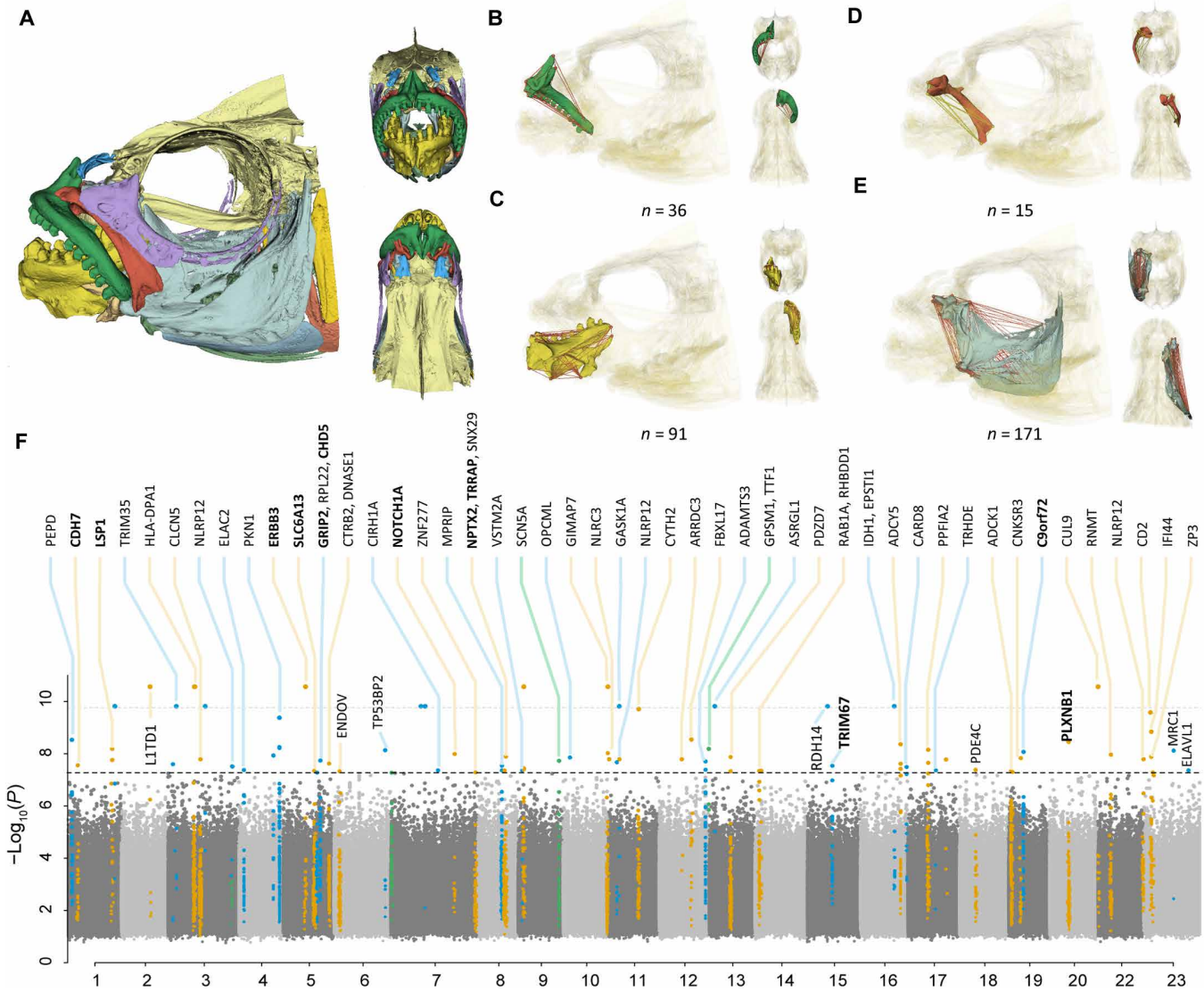

Fig. 4. Genotype-phenotype mapping of the head asymmetries. (**A**) Craniofacial structure in lateral, front, and dorsal views with each color representing a separate ossified unit. (**B**) A total of 36 linear distance traits on the premaxilla is displayed in lateral, front, and dorsal views. (**C**) A total of 91 linear distance traits on the mandible is displayed in the lateral, front, and dorsal views. (**D**) A total of 15 linear distance traits on the maxilla is displayed in the lateral, front, and dorsal views. (**E**) A total of 171 linear distance traits on the suspensorium is displayed in the lateral, front, and dorsal views. (**F**) Aggregate of the GWAS signals detected (across all traits). The chromosomal region (and surrounding genes) showing the strongest association to a trait is indicated above each GWAS peak. The genes mentioned in the main text are bolded. The two horizontal dashed lines represent the genome-wide significant variants (P value $<5 \times 10^{-8}$) and the study-wide significance (P value $<1.58 \times 10^{-10}$).

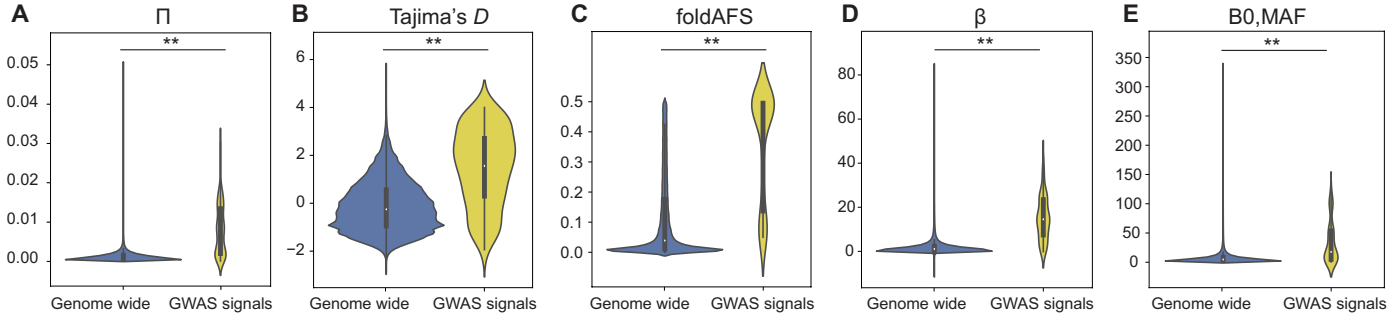

Fig. 5. Genetic characteristic comparison between genome-wide variants and candidate GWAS signals. (**A**) Comparison of Π between genome-wide variants and candidate GWAS signals. A t test assessed significance, with double asterisks indicating a P value <0.05. (**B**) Comparison of Tajima's D between genome-wide variants and candidate GWAS signals. (**C**) Comparison of foldAFS between genome-wide variants and candidate GWAS signals. (**D**) Comparison of β between genome-wide variants and candidate GWAS signals. (**E**) Comparison of B0,MAF between genome-wide variants and candidate GWAS signals.

Downloaded from https://www.science.org on July 31, 2025

signature of balancing selection in this species. High-quality μCT data from hundreds of individuals detailed the craniofacial changes and revealed a solid bimodal distribution of head asymmetry in this species. Our GWAS analyses identified 72 loci in the genome and enriched genes involved in left-right asymmetry establishment, facial morphology, and the nervous system.

As we found in this study and before (*21*), the lateral head asymmetry of this species fits a bimodal distribution, manifested as bimodal distributions of overall volume asymmetries and a large number of linear asymmetries in detail. On the basis of prior work on polymorphism in other species of cichlid fish, one might have expected that also here, the underlying genetic basis is due to a small number of loci with large effects; we had identified that single loci determine the gold and dark morphs in Midas cichlids as well as their lip sizes (*61*, *62*). However, in this study, we instead find that a relatively large number of small effect loci seem to contribute to this symmetry-breaking head polymorphism and lateralized feeding behavior. This indicates that head asymmetry is not an independent trait but is closely linked to lateralized feeding behavior, which might precede the morphological asymmetry during the fish's ontogeny. There is some evidence that wild individuals exhibit more pronounced asymmetry compared to laboratory-reared fish (*22*), and previous research suggested that one-side feeding treatment can induce certain morphological changes (*19*). Therefore, the interaction between early lateralized behavior and subsequent morphological asymmetry may combine to lead to the final bimodal asymmetry. Consistent with this, the genes we identified in GWAS are not only related to craniofacial development (e.g., *LSP1*, *PLXNB1*, and *ERRBB3*), such as osteocyte development, epithelial cell migration, etc., but we also identified a certain number of nervous system–related genes (e.g., *SLC6A13*, *TRRAP*, *TRIM67*, *CHD5*, *GRIP2*, and *NPTX2*) that might contribute to lateralized behavioral preferences. This is reminiscent of a similar hypothesis regarding handedness in humans and the promotion of muscle strength (for instance, muscle type proportion), a difference that is exacerbated by the frequency of hand use (*63*, *64*) and in turn influences its hand preference (*65*).

Our analysis points toward two genes that might play a role in breaking symmetry, *notch1a* and cadherin-7, which are known to be highly associated with the early establishment of body asymmetry in animals. *Notch* activity is well known for its functions in inducing left-sided *Nodal* expression in the lateral plate mesoderm, a conserved mechanism essential for establishing normal left-right asymmetry in vertebrate embryos (*38*, *66*). Cadherins are a diverse group of cell adhesion molecules that regulate many morphogenetic events during animal development (*67*). Disruption of the function of N-cadherin during chick gastrulation leads to a random orientation of the heart along the left-right axis (*42*). Together, the most likely interpretation for the identified genetic basis is that left-right asymmetry arises from the activation of upstream asymmetry signals during early embryonic development in the lateral plate mesoderm, combined with the interplay of downstream genes related to facial morphology and neural development, as well as environmental factors (such as experience). Together, these elements contribute to the asymmetric polymorphism observed in the fish population. Although the identified regulators of asymmetry (e.g., notch family members) are intriguing, we did not observe broader asymmetries beyond head morphology and behavior. Therefore, further investigation of these genes is warranted.

Our findings at the genomic level support the assumed role of balancing selection from ecological theory (*14*, *15*). The genetic signature found is in agreement with one that would be expected to be associated with the maintenance of adaptive genetic diversity within the genetic regions responsible for bilateral head asymmetry. This also agrees with predictions of antagonistic predator-prey interaction (*14*, *68*) and negative frequency–dependent selection as a special form of balancing selection, suggesting that adaptive evolution can result in a genetically diverse set of segregating alleles rather than a single optimal genotype (*69*, *70*).

In conclusion, our study has identified a largely multigenic basis for the symmetry-breaking head and feeding morphs (in particular 72 genomic regions) influencing the bilateral head asymmetry in *P. microlepis*. These genomic regions harbor strongly suggestive candidate genes, whose functions combined implicate craniofacial and behavioral basis that might affect the origin and maintenance of these feeding morphs. Knowing the genetic basis of such complex traits will not only enrich knowledge on how the symmetric head morphology that dominates the animal kingdom has evolved, been retained, and occasionally broken but also potentially help to unveil congenital causes of similar human pathological disorders, such as craniofacial malformations and neurological disorders.

MATERIALS AND METHODS

Genome sequencing and assembly

The genome sample was extracted from the liver tissue of an adult female specimen collected from Lake Tanganyika, Zambia, and preserved in the animal research facility (TFA) at the University of Konstanz. This work was carried out following the guidelines and regulations of the Animal Care and Ethics Committee of Germany.

Extraction of high-molecular-weight genomic DNA

The high-molecular-weight (HMW) genomic DNA (gDNA) of *P. microlepis* was extracted with the Circulomics Nanobind Tissue Big DNA kit according to the manufacturer's instructions. Briefly, liver tissue was minced into small slices on a clean and cold surface and lastly homogenized with a TissueRuptor II device (Qiagen). After complete tissue lysis with proteinase K, the remaining cell debris was removed, and the gDNA was bound to the Circulomics Nanobind disk upon addition of salting buffer and isopropanol. HMW gDNA was eluted from the nanobind disks in the elution buffer.

PacBio HiFi library preparation and sequencing

The PacBio HiFi library of extracted gDNA of *P. microlepis* was prepared as recommended by Pacific Biosciences according to the use of the SMRTbell Express Template Prep Kit 2.0. In summary, HMW gDNA was sheared on the MegaRuptor device (Diagenode) and then used for library preparation. Size selection was performed for fragments larger than 5.5 kb with the BluePippin device (Fragment Analyzer, Agilent Technologies). The size-selected library ran on one Sequel II SMRT cell for 30 hours.

Hi-C chromatin conformation capture

Chromatin conformation capture was conducted using the ARIMA Hi-C+ Kit. A small amount of flash-frozen tissue was cross-linked, and the resulting gDNA was digested with a four-enzyme cocktail. The 5′-overhangs were biotin labeled, and spatially proximal DNA fragments were ligated. These fragments were enriched for Illumina library preparation using the KAPA HyperPrep kit. The barcoded Hi-C library was sequenced on a NovaSeq6000 S4 flow cell with 2 × 150 cycles.

Genome assembly

Circular consensus (HiFi) reads were generated from the raw subreads using PacBio's CCS module, requiring a minimum of three

Downloaded from https://www.science.org on July 31, 2025

passes. These reads were then refined with DeepConsensus by mapping subreads to CCS reads and processing the resulting bam files. The remaining SMRTbell adapters were trimmed using Cutadapt. Hi-C reads were quality controlled and trimmed to a minimum phred-quality score of 20 using TrimGalore.

The genome assembly was performed following the Vertebrate Genomes Project pipeline. The PacBio HiFi reads were assembled into two haplotype assemblies of contigs using the Hi-C reads and hifiasm (options: -l 3 --h1 --h2) (*71*). Purge Dups was used to remove residual haplotigs. Hi-C reads were aligned to hap1 and hap2 assemblies via the Arima mapping pipeline, and scaffolding was completed with YaHS (*72*).

Contigs and scaffolds were checked for contaminants using the NCBI Foreign Contamination Screen, with no contamination detected. Manual curation was done with the RAPID pipeline, realigning sequencing data and visualizing Hi-C heatmaps via HiGlass (*73*). Hi-C contact maps confirmed the successful chromosome-scale scaffolding of both haplotypes.

Genome annotation

Repeat sequences

The repetitive sequences of the two final assemblies were annotated using both de novo and homology-based methods. First, RepeatModeler version 2.0.3 (*74*) was used to construct a de novo transposable element library with default parameters. RepeatModeler uses three de novo repeat finding programs—RECON version 1.0.8 (*75*), RepeatScout version 1.0.6 (*76*), and LTR_retriever version 2.9.0 (*77*)—to identify repeat element boundaries and family relationships from DNA sequences. Then, RepeatMasker version 4.1.2-p1 (www.repeatmasker.org) was used to screen interspersed repeats and low-complexity DNA sequences in the genome with the default transposable element Dfam database and the de novo repeat library constructed using RepeatModeler. As a complement, tandem repeats were predicted using TRF version 4.09 (*76*).

Annotation of protein-coding genes

Protein-coding genes were predicted by combining transcriptome evidence with homology-based evidence aligned to the genome. For homology-based evidence, we first collected the protein sequences of the vertebrata from OrthoDB version 11 (*78*). We also downloaded all the protein sequences of Cichliformes from NCBI (date: 21 November 2023). Then, the two sets of protein sequences were combined as the final protein evidence. For the transcriptome evidence, we collected the RNA sequencing data from the previous studies (*32*, *79*). With this evidence, braker3 version 3.0.6 (*80*) was implemented to perform the subsequent gene annotation. In detail, the RNA sequencing data were aligned to the genome by HISAT2 version 2.2.1 (*81*) and then assembled with StringTie2 version 2.2.1 (*82*). The assembled transcripts and the protein evidence were analyzed by GeneMark-ETP version 1.02 (*83*) to output the high-confidence genes. Augustus version 3.5.0 (*84*) was trained using the set of high-confidence genes and then was run to do gene prediction. Then, TSEBRA (*85*) was used to combine the predictions made by AUGUSTUS and GeneMark-ETP while integrating the high-confidence genes directly into the result to ensure their inclusion. Last, the functional annotation of the final gene set was performed by the InterProscan5 (*86*) and Eggnog-Mapper (*87*) pipelines to assign gene symbols and names. Protein sequences were clustered into paralogous and orthologous sequences using OrthoFinder (*88*) with default parameters.

Population sampling, raw read processing, and SNP calling

A total of 102 wild-caught *P. microlepis* specimens was collected from five locations in Zambia (table S1) in April 2018, following local regulations and guidelines (Permit G. R. no.: 2077761) (*17*). All the freshly caught specimens were preserved in 98% ethanol and stored at 4°C for further studies.

HMW DNA was extracted from muscle tissue from all 102 specimens using commercial kits (Genaxxon Tissue DNA Purification kit) without a ribonuclease A treatment step. DNA integrity was inspected using agarose gel electrophoresis on Agilent Tape Station 4150, and concentrations were determined on a QuBit4.0 fluorometer. Genomic libraries were prepared on the Illumina platform and were paired-end sequenced (2 * 150 base pairs) on the DNBSEQ platform at the Beijing Genomics Institute (BGI, Hong Kong).

Raw reads were trimmed using Trimmomatic version 0.39 (*89*) to remove low-quality bases and adapter sequences. Cleaned reads from all individuals were aligned to the *P. microlepis* genome generated in this study using the bwa "BWA-MEM" algorithm (BWA 0.7.17-r1188) (*90*) with default parameters. On average, 99.6% of the reads were mapped, resulting in a final average sequencing coverage of 27.6× (ranging from 23.3× to 28.2×) per individual. Alignment bam/sam files were next subjected to a series of processing and filtration including position sorting, duplicated read marking and removal, local realignment, and base quality recalibration, which were carried out using tools available in Picard (version 2.26.9-0) and Genome Analysis Toolkit (GATK version 4.2.5.0) packages (*91*). SNPs were genotyped using UnifiedGenotyper in the GATK package, and the output was further filtered using VCFtools (version 0.1.16) (*92*). SNPs that did not meet the following criteria were excluded: (i) greater than one-third of corresponding mean sequencing depth and <3 * corresponding mean sequencing depth; (ii) a minor allele frequency >0.05; (iii) maximum missing rate <0.1; (iv) only biallelic sites; and (v) minimum quality >30. In total, 2,727,283 SNPs across the 102 samples passed these filters. A principal components analysis was performed to visualize the population structure using the whole-genome SNPs with PLINK (version 1.90b6.21) (*93*).

3D μCT and head asymmetry quantification

The 102 ethanol-preserved specimens were transferred to 4% formalin for μCT scanning. The μCT scans were conducted at the University of Ulm using a Skyscan 1176 (Bruker, Kontich, Belgium). The detailed parameters of the μCT setting refer to Tian *et al.* (*21*). The resulting images were reconstructed using NRecon software and processed with 3D Slicer (*94*).

Three craniofacial skeletal elements, the premaxilla, maxilla, and mandible, which directly contact prey during scale eating, were isolated to measure the skeleton volume. The volume asymmetry of each skeleton was calculated on the basis of the equation "Degree of asymmetry = (Left − Right)/(Left + Right)" (*95*). Shape variations were quantified with 3D geometric morphometrics analysis, which involved placing 96 landmarks on homologous and anatomically recognizable features on the head (fig. S4). The selection of bones and the placement of landmarks were guided by their known biological functions and prior practical experience (*21*). Major 3D shape variations among the fish were extracted using generalized Procrustes analysis (*96*, *97*) and then summarized with principal components analysis (*98*). A total of 313 linear distance asymmetries was measured by comparing the difference in linear distance (physical

Downloaded from https://www.science.org on July 31, 2025

distance between two 3D landmarks) between the left and right sides using the same equation above.

Genome-wide association analysis

Before GWAS, the variants with more than 10% missing data or with a MAF of less than 0.05 were excluded. To identify genome regions associated with head asymmetry variation in *P. microlepis*, we performed the linear mixed model approach implemented in GEMMA (version 0.98.3) (*99*) using the -lmm 4 option. PLINK (version 1.90b6.21) (*93*) was used to prepare the input files (genotype and phenotype) for GEMMA. In GEMMA, relatedness was accounted for by including a centered kinship matrix as a random effect. Association was tested for each of the 313 morphological traits and then combined for visualization. Variations were only considered significant if the *P* value of the Wald test was smaller than the conventional threshold (5×10^{-8}). PLINK was used to identify the SNPs within the same linkage block with a candidate SNP with the parameter settings --ld-window-r2 0.2 and --ld-window-kb 1000.

Detecting genomic signatures of balancing selection

To test the hypothesis that balancing selection plays a role in *P. microlepis*, genomic scans were performed using BetaScan (β) (*33*) and BalLeRMiX (B0,MAF) (*34*). BetaScan relies on summary statistics to detect correlations in allele frequencies between core-balanced SNPs and neighboring neutral SNPs. High β scores indicate an excess of SNPs at similar frequencies. The other method, B0,MAF scores, is based on the composition-likelihood ratio test statistic, which distinguishes departures from neutral evolution. Higher B0,MAF indicates a larger possibility of favoring balancing selection and rejecting genome-wide empirical distribution (neutrality).

Gene set enrichment analysis

We characterized the potential functions of genes of interest by extracting the protein sequences from the newly generated *P. microlepis* reference and using these sequences to search for overrepresented GO terms using hypergeometric tests as implemented in the KOBAS version 3.0 server (*100*). Benjamini-Hochberg false discovery rate correction was performed to adjust for multiple tests. GO terms with a false discovery rate–corrected *P* value <0.01 were considered significantly enriched. Furthermore, we searched for the genes of interest within a curated list of fish genes to further investigate their functions and associated pathways, particularly in the context of their roles in fish (*101*).

Supplementary Materials

The PDF file includes:

Figs. S1 to S12
Tables S1 to S6
Legends for data S1 to S7

Other Supplementary Material for this manuscript includes the following:

Data S1 to S7

REFERENCES AND NOTES

1. C. Darwin, *The Origin of Species* (PF Collier & Son, 1909).
2. E. B. Ford, Polymorphism. *Biol. Rev. Camb. Philos. Soc.* **20**, 73–88 (1945).
3. S. Ni, G. A. McCulloch, G. C. Kroos, T. M. King, L. Dutoit, B. J. Foster, K. Hema, J. M. Jandt, M. Peng, P. K. Dearden, J. M. Waters, Human-driven evolution of color in a stonefly mimic. *Science* **386**, 453–458 (2024).
4. P. Jay, M. Chouteau, A. Whibley, H. Bastide, H. Parrinello, V. Llaurens, M. Joron, Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat. Genet.* **53**, 288–293 (2021).
5. G. A. Jamie, J. I. Meier, The persistence of polymorphisms across species radiations. *Trends Ecol. Evol.* **35**, 795–808 (2020).
6. R. Arbore, S. Barbosa, J. Brejcha, Y. Ogawa, Y. Liu, M. P. J. Nicolai, P. Pereira, S. J. Sabatino, A. Cloutier, E. S. K. Poon, C. I. Marques, P. Andrade, G. Debruyn, S. Afonso, R. Afonso, S. G. Roy, U. Abdu, R. J. Lopes, P. Mojzes, P. Marsik, S. Y. W. Sin, M. A. White, P. M. Araujo, J. C. Corbo, M. Carneiro, A molecular mechanism for bright color variation in parrots. *Science* **386**, eadp7710 (2024).
7. M. J. Wittmann, A. O. Bergland, M. W. Feldman, P. S. Schmidt, D. A. Petrov, Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9932–E9941 (2017).
8. E. C. Moore, P. J. Ciccotto, E. N. Peterson, M. S. Lamm, R. C. Albertson, R. B. Roberts, Polygenic sex determination produces modular sex polymorphism in an African cichlid fish. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2118574119 (2022).
9. R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974), vol. 560.
10. M. B. Hamilton, *Population Genetics* (John Wiley & Sons, 2021).
11. D. Charlesworth, Balancing selection and its effects on sequences in nearby genome regions. *PLOS Genet.* **2**, e64 (2006).
12. F. M. Key, J. C. Teixeira, C. de Filippo, A. M. Andres, Advantageous diversity maintained by balancing selection in humans. *Curr. Opin. Genet. Dev.* **29**, 45–51 (2014).
13. L. D. Gigord, M. R. Macnair, A. Smithson, Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soo. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6253–6255 (2001).
14. M. Hori, Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science* **260**, 216–219 (1993).
15. D. Futuyma, *Evolution* (Sinauer Associates, 2009).
16. H. J. Lee, S. Pittlik, J. C. Jones, W. Salzburger, M. Barluenga, A. Meyer, Genetic support for random mating between left and right-mouth morphs in the dimorphic scale-eating cichlid fish *Perissodus microlepis* from Lake Tanganyika. *J. Fish Biol.* **76**, 1940–1957 (2010).
17. H. Kusche, H. J. Lee, A. Meyer, Mouth asymmetry in the textbook example of scale-eating cichlid fish is not a discrete dimorphism after all. *Proc. Biol. Sci.* **279**, 4715–4723 (2012).
18. T. A. Stewart, R. C. Albertson, Evolution of a unique predatory feeding apparatus: Functional anatomy, development and a genetic locus for jaw laterality in Lake Tanganyika scale-eating cichlids. *BMC Biol.* **8**, 8 (2010).
19. T. J. Van Dooren, H. A. van Goor, M. van Putten, Handedness and asymmetry in scale-eating cichlids: Antisymmetries of different strength. *Evolution* **64**, 2159–2165 (2010).
20. R. Takahashi, K. Watanabe, M. Nishida, M. Hori, Evolution of feeding specialization in Tanganyikan scale-eating cichlids: A molecular phylogenetic approach. *BMC Evol. Biol.* **7**, 195 (2007).
21. X. Tian, S. Lee, J. Tuckermann, A. Meyer, Bilateral asymmetry in craniofacial structures and kinematics of feeding attacks in the scale-eating cichlid fish, *Perissodus microlepis*. *Zool. Res.* **46**, 370–378 (2024).
22. H. J. Lee, H. Kusche, A. Meyer, Handed foraging behavior in scale-eating cichlid fish: Its potential role in shaping morphological asymmetry. *PLOS ONE* **7**, e44670 (2012).
23. Y. Takeuchi, M. Hori, S. Tada, Y. Oda, Acquisition of lateralized predation behavior associated with development of mouth asymmetry in a lake tanganyika scale-eating cichlid fish. *PLOS ONE* **11**, e0147476 (2016).
24. F. Berlinghieri, P. Panizzon, I. L. Penry-Williams, C. Brown, Laterality and fish welfare - A review. *Appl. Anim. Behav. Sci.* **236**, 105239 (2021).
25. F. Raffini, A. Meyer, A comprehensive overview of the developmental basis and adaptive significance of a textbook polymorphism: Head asymmetry in the cichlid fish *Perissodus microlepis*. *Hydrobiologia* **832**, 65–84 (2019).
26. M. Hori, H. Ochi, M. Kohda, Inheritance pattern of lateral dimorphism in two cichlids (a scale eater, *Perissodus microlepis*, and an herbivore, *Neolamprologus moorii*) in Lake Tanganyika. *Zoolog. Sci.* **24**, 486–492 (2007).
27. H. Hata, M. Yasugi, Y. Takeuchi, S. Takahashi, M. Hori, Measuring and evaluating morphological asymmetry in fish: Distinct lateral dimorphism in the jaws of scale-eating cichlids. *Ecol. Evol.* **3**, 4641–4647 (2013).
28. F. Raffini, C. Fruciano, P. Franchini, A. Meyer, Towards understanding the genetic basis of mouth asymmetry in the scale-eating cichlid *Perissodus microlepis*. *Mol. Ecol.* **26**, 77–91 (2017).
29. F. Raffini, C. Fruciano, A. Meyer, Morphological and genetic correlates in the left–right asymmetric scale-eating cichlid fish of Lake Tanganyika. *Biol. J. Linn. Soc.* **124**, 67–84 (2018).
30. H. J. Lee, V. Heim, A. Meyer, Genetic and environmental effects on the morphological asymmetry in the scale-eating cichlid fish, *Perissodus microlepis*. *Ecol. Evol.* **5**, 4277–4286 (2015).
31. Y. Takeuchi, Y. Higuchi, K. Ikeya, M. Tagami, Y. Oda, Experience-dependent learning of behavioral laterality in the scale-eating cichlid *Perissodus microlepis* occurs during the early developmental stage. *Sci. Rep.* **12**, 723 (2022).

Downloaded from https://www.science.org on July 31, 2025

32. H. J. Lee, R. F. Schneider, T. Manousaki, J. H. Kang, E. Lein, P. Franchini, A. Meyer, Lateralized feeding behavior is associated with asymmetrical neuroanatomy and lateralized gene expressions in the brain in scale-eating cichlid fish. *Genome Biol. Evol.* **9**, 3122–3136 (2017).
33. K. M. Siewert, B. F. Voight, Detecting long-term balancing selection using allele frequency correlation. *Mol. Biol. Evol.* **34**, 2996–3005 (2017).
34. X. Cheng, M. DeGiorgio, Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *Mol. Biol. Evol.* **37**, 3267–3291 (2020).
35. M. Bradl, H. Lassmann, Oligodendrocytes: Biology and pathology. *Acta Neuropathol.* **119**, 37–53 (2010).
36. W. J. Tang, J. G. Fernandez, J. J. Sohn, C. T. Amemiya, Chitin is endogenously produced in vertebrates. *Curr. Biol.* **25**, 897–900 (2015).
37. Y.-H. Chen, H. J. P. Zhao, Evolution of digestive enzymes and dietary diversification in birds. *PeerJ* **7**, e6840 (2019).
38. L. T. Krebs, N. Iwai, S. Nonaka, I. C. Welsh, Y. Lan, R. Jiang, Y. Saijoh, T. P. O'Brien, H. Hamada, T. Gridley, Notch signaling regulates left–right asymmetry determination by inducing *Nodal* expression. *Genes Dev.* **17**, 1207–1212 (2003).
39. N. Kishimoto, K. Asakawa, R. Madelaine, P. Blader, K. Kawakami, K. Sawamoto, Interhemispheric asymmetry of olfactory input-dependent neuronal specification in the adult brain. *Nat. Neurosci.* **16**, 884–888 (2013).
40. V. F. Naturale, M. A. Pickett, J. L. Feldman, E-cadherin/*HMR-1* and *PAR-3* break symmetry at stable cell contacts in a developing epithelium. bioRxiv 503536 [Preprint] (2022).
41. A. González-Reyes, D. St. Johnston, The *Drosophila* AP axis is polarised by the cadherin-mediated positioning of the oocyte. *Development* **125**, 3635–3644 (1998).
42. M. I. García-Castro, E. Vielmetter, M. Bronner-Fraser, N-Cadherin, a cell adhesion molecule involved in establishment of embryonic left-right asymmetry. *Science* **288**, 1047–1051 (2000).
43. B. Bonfante, P. Faux, N. Navarro, J. Mendoza-Revilla, M. Dubied, C. Montillot, E. Wentworth, L. Poloni, C. Varón-González, P. Jones, Z. Xiong, M. Fuentes-Guajardo, S. Palmal, J. C. Chacón-Duque, M. Hurtado, V. Villegas, V. Granja, C. Jaramillo, W. Arias, R. Barquera, P. Everardo-Martínez, M. Sánchez-Quinto, J. Gómez-Valdés, H. Villamil-Ramírez, C. C. S. de Cerqueira, T. Hünemeier, V. Ramallo, F. Liu, S. M. Weinberg, J. R. Shaffer, E. Stergiakouli, L. J. Howe, P. G. Hysi, T. D. Spector, R. Gonzalez-José, L. Schüler-Faccini, M.-C. Bortolini, V. Acuña-Alonzo, S. Canizales-Quinteros, C. Gallo, G. Poletti, G. Bedoya, F. Rothhammer, C. Thauvin-Robinet, L. Faivre, C. Costedoat, D. Balding, T. Cox, M. Kayser, L. Duplomb, B. Yalcin, J. Cotney, K. Adhikari, A. Ruiz-Linares, A GWAS in Latin Americans identifies novel face shape loci, implicating *VPS13B* and a Denisovan introgressed region in facial variation. *Sci. Adv.* **7**, eabc6160 (2021).
44. L. Fagerberg, B. M. Hallström, P. Oksvold, C. Kampf, D. Djureinovic, J. Odeberg, M. Habuka, S. Tahmasebpoor, A. Danielsson, K. Edlund, A. Asplund, E. Sjöstedt, E. Lundberg, C. A.-K. Szigyarto, M. Skogs, J. O. Takanen, H. Berling, H. Tegel, J. Mulder, P. Nilsson, J. M. Schwenk, C. Lindskog, F. Danielsson, A. Mardinoglu, A. Sivertsson, K. von Feilitzen, M. Forsberg, M. Zwahlen, I. Olsson, S. Navani, M. Huss, J. Nielsen, F. Ponten, M. Uhlén, Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).
45. M. M. Hossain, "Phosphorylation and sumoylation help maintain the functions of endothelial cell-expressed leukocyte-specific protein 1 (*LSP1*)," thesis, University of Saskatchewan (2016).
46. N. Wang, X. Liu, Z. Tang, X. Wei, H. Dong, Y. Liu, H. Wu, Z. Wu, X. Li, X. Ma, Z. Guo, Increased BMSC exosomal *miR-140-3p* alleviates bone degradation and promotes bone restoration by targeting *Plxnb1* in diabetic rats. *J. Nanotechnol.* **20**, 97 (2022).
47. T. Negishi-Koga, M. Shinohara, N. Komatsu, H. Bito, T. Kodama, R. H. Friedel, H. Takayanagi, Suppression of bone formation by osteoclastic expression of semaphorin 4D. *Nat. Med.* **17**, 1473–1480 (2011).
48. Q. Li, P. Xu, C. Zhang, Y. Gao, MiR-362-5p inhibits cartilage repair in osteoarthritis via targeting plexin B1. *J. Orthop. Surg. (Hong Kong)* **30**, 10225536221139887 (2022).
49. K. Fried, M. Risling, H. Tidcombe, M. Gassmann, C. Lillesaar, Expression of ErbB3, ErbB4, and neuregulin-1 mRNA during tooth development. *Dev. Dyn.* **224**, 356–360 (2002).
50. B. Cogné, S. Ehresmann, E. Beauregard-Lacroix, J. Rousseau, T. Besnard, T. Garcia, S. Petrovski, S. Avni, K. McWalter, P. R. Blackburn, S. J. Sanders, K. Uguen, J. Harris, J. S. Cohen, M. Blyth, A. Lehman, J. Berg, M. H. Li, U. Kini, S. Joss, C. von der Lippe, C. T. Gordon, J. B. Humberson, L. Robak, D. A. Scott, V. R. Sutton, C. M. Skraban, J. J. Johnston, A. Poduri, M. Nordenskjöld, V. Shashi, E. H. Gerkes, E. M. H. F. Bongers, C. Gilissen, Y. A. Zarate, M. Kvarnung, K. P. Lally, P. A. Kulch, B. Daniels, A. Hernandez-Garcia, N. Stong, J. McGaughran, K. Retterer, K. Tveten, J. Sullivan, M. R. Geisheker, A. Stray-Pedersen, J. M. Tarpinian, E. W. Klee, J. C. Sapp, J. Zyskind, Ø. L. Holla, E. Bedoukian, F. Filippini, A. Guimier, A. Picard, Ø. L. Busk, J. Punetha, R. Pfundt, A. Lindstrand, A. Nordgren, F. Kalb, M. Desai, A. H. Ebanks, S. N. Jhangiani, T. Dewan, Z. H. C. Akdemir, A. Telegrafi, E. H. Zackai, A. Begtrup, X. Song, A. Toutain, I. M. Wentzensen, S. Odent, D. Bonneau, X. Latypova, W. Deb, CAUSES Study, S. Redon, F. Bilan, M. Legendre, C. Troyer, K. Whitlock, O. Caluseriu, M. I. Murphree, P. N. Pichurin, K. Agre, R. Gavrilova, T. Rinne, M. Park, C. Shain, E. L. Heinzen, R. Xiao, J. Amiel, S. Lyonnet, B. Isidor, L. G. Biesecker, D. Lowenstein, J. E. Posey, A.-S. Denommé-Pichon, Deciphering Developmental Disorders study, C. Férec, X.-J. Yang, J. A. Rosenfeld, B. Gilbert-Dussardier, S. Audebert-Bellanger, R. Redon, H. A. F. Stessman, C. Nellaker, Y. Yang, J. R. Lupski, D. B. Goldstein, E. E. Eichler, F. Bolduc, S. Bézieau, S. Küry, P. M. Campeau, Missense variants in the histone acetyltransferase complex component gene *TRRAP* cause autism and syndromic intellectual disability. *Am. J. Hum. Genet.* **104**, 530–541 (2019).
51. A. Soldan, S. Oh, T. Ryu, C. Pettigrew, Y. Zhu, A. Moghekar, M. F. Xiao, G. M. Pontone, M. Albert, C. H. Na, P. Worley, *NPTX2* in cerebrospinal fluid predicts the progression from normal cognition to mild cognitive impairment. *Ann. Neurol.* **94**, 620–631 (2023).
52. V. Babenko, D. Smagin, I. Kovalenko, A. Galyamina, N. N. Kudryavtseva, Differentially expressed genes of the *Slc6a* family as markers of altered brain neurotransmitter system function in pathological states in mice. *Neurosci. Behav. Physiol.* **50**, 199–209 (2020).
53. H. Yaguchi, F. Okumura, H. Takahashi, T. Kano, H. Kameda, M. Uchigashima, S. Tanaka, M. Watanabe, H. Sasaki, S. Hatakeyama, *TRIM67* protein negatively regulates Ras activity through degradation of 80K-H and induces neuritogenesis. *J. Biol. Chem.* **287**, 12050–12059 (2012).
54. E. M. Devenney, R. M. Ahmed, G. Halliday, O. Piguet, M. C. Kiernan, J. R. Hodges, Psychiatric disorders in *C9orf72* kindreds: Study of 1,414 family members. *Neurology* **91**, e1498–e1507 (2018).
55. I. Parenti, D. Lehalle, C. Nava, E. Torti, E. Leitão, R. Person, T. Mizuguchi, N. Matsumoto, M. Kato, K. Nakamura, S. A. de Man, H. Cope, V. Shashi, Undiagnosed Diseases Network, J. Friedman, P. Joset, K. Steindl, A. Rauch, I. Muffels, P. M. van Hasselt, F. Petit, T. Smol, G. Le Guyader, F. Bilan, A. Sorlin, A. Vitobello, C. Philippe, I. M. B. H. van de Laar, M. A. van Slegtenhorst, P. M. Campeau, P. Y. B. Au, M. Nakashima, H. Saitsu, T. Yamamoto, Y. Nomura, R. J. Louie, M. J. Lyons, A. Dobson, A. S. Plomp, M. M. Motazacker, F. J. Kaiser, A. T. Timberlake, S. A. Fuchs, C. Depienne, C. Mignot, Missense and truncating variants in *CHD5* in a dominant neurodevelopmental disorder with intellectual disability, behavioral disturbances, and epilepsy. *Hum. Genet.* **140**, 1109–1120 (2021).
56. W. Lu, E. B. Ziff, *PICK1* interacts with ABP/*GRIP* to regulate AMPA receptor trafficking. *Neuron* **47**, 407–421 (2005).
57. N. Nguyen, "Molecular basis of lateralization in chick brain," thesis, Macquarie University (2022).
58. G. Cuellar-Partida, J. Y. Tung, N. Eriksson, E. Albrecht, F. Aliev, O. A. Andreassen, I. Barroso, J. S. Beckmann, M. P. Boks, D. I. Boomsma, H. A. Boyd, M. M. B. Breteler, H. Campbell, D. I. Chasman, L. F. Cherkas, G. Davies, E. J. C. de Geus, I. J. Deary, P. Deloukas, D. M. Dick, D. L. Duffy, J. G. Eriksson, T. Esko, B. Feenstra, F. Geller, C. Gieger, I. Giegling, S. D. Gordon, J. Han, T. F. Hansen, A. M. Hartmann, C. Hayward, K. Heikkila, A. A. Hicks, J. N. Hirschhorn, J. J. Hottenga, J. E. Huffman, L. D. Hwang, M. A. Ikram, J. Kaprio, J. P. Kemp, K. T. Khaw, N. Klopp, B. Konte, Z. Kutalik, J. Lahti, X. Li, R. J. F. Loos, M. Luciano, S. H. Magnusson, M. Mangino, P. Marques-Vidal, N. G. Martin, W. L. McArdle, M. I. McCarthy, C. Medina-Gomez, M. Melbye, S. A. Melville, A. Metspalu, L. Milani, V. Mooser, M. Nelis, D. R. Nyholt, K. S. O'Connell, R. A. Ophoff, C. Palmer, A. Palotie, T. Palviainen, G. Pare, L. Paternoster, L. Peltonen, B. Penninx, O. Polasek, P. P. Pramstaller, I. Prokopenko, K. Raikkonen, S. Ripatti, F. Rivadeneira, I. Rudan, D. Rujescu, J. H. Smit, G. D. Smith, J. W. Smoller, N. Soranzo, T. D. Spector, B. S. Pourcain, J. M. Starr, H. Stefansson, S. Steinberg, M. Teder-Laving, G. Thorleifsson, K. Stefansson, N. J. Timpson, A. G. Uitterlinden, C. M. van Duijn, F. J. A. van Rooij, J. M. Vink, P. Vollenweider, E. Vuoksimaa, G. Waeber, N. J. Wareham, N. Warrington, D. Waterworth, T. Werge, H. E. Wichmann, E. Widen, G. Willemsen, A. F. Wright, M. J. Wright, M. Xu, J. H. Zhao, P. Kraft, D. A. Hinds, C. M. Lindgren, R. Magi, B. M. Neale, D. M. Evans, S. E. Medland, Genome-wide association study identifies 48 common genetic variants associated with handedness. *Nat. Hum. Behav.* **5**, 59–70 (2021).
59. Z. Sha, D. Schijven, A. Carrion-Castillo, M. Joliot, B. Mazoyer, S. E. Fisher, F. Crivello, C. Francks, The genetic architecture of structural left-right asymmetry of the human brain. *Nat. Hum. Behav.* **5**, 1226–1239 (2021).
60. T. Takahashi, M. Hori, Evidence of disassortative mating in a Tanganyikan cichlid fish and its role in the maintenance of intrapopulation dimorphism. *Biol. Lett.* **4**, 497–499 (2008).
61. A. F. Kautt, C. F. Kratochwil, A. Nater, G. Machado-Schiaffino, M. Olave, F. Henning, J. Torres-Dowdall, A. Harer, C. D. Hulsey, P. Franchini, M. Pippel, E. W. Myers, A. Meyer, Contrasting signatures of genomic divergence during sympatric speciation. *Nature* **588**, 106–111 (2020).
62. C. F. Kratochwil, A. F. Kautt, A. Nater, A. Harer, Y. Liang, F. Henning, A. Meyer, An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes. *Nat. Commun.* **13**, 296 (2022).
63. A. R. Fugl-meyer, A. Eriksson, M. Sjöström, G. Söderström, Is muscle structure influenced by genetical or functional factors? A study of three forearm muscles. *Acta Physiol. Scand.* **114**, 277–281 (1982).
64. A. Adam, C. J. De Luca, Z. Erim, Hand dominance and motor unit firing behavior. *J. Neurophysiol.* **80**, 1373–1382 (1998).

Downloaded from https://www.science.org on July 31, 2025

65. R. L. McGrath, S. S. Kantak, Reduced asymmetry in motor skill learning in left-handed compared to right-handed individuals. *Hum. Mov. Sci.* **45**, 130–141 (2016).
66. A. Raya, Y. Kawakami, C. Rodrıguez-Esteban, D. Büscher, C. M. Koth, T. Itoh, M. Morita, R. M. Raya, I. Dubova, J. G. J. G. Bessa, J. L. de la Pompa, J. C. I. Belmonte, Notch activity induces *Nodal* expression and mediates the establishment of left–right asymmetry in vertebrate embryos. *Genes Dev.* **17**, 1213–1218 (2003).
67. U. Tepass, K. Truong, D. Godt, M. Ikura, M. Peifer, Cadherins in embryonic and neural morphogenesis. *Nat. Rev. Mol. Cell Biol.* **1**, 91–100 (2000).
68. H. Ichijo, T. Nakamura, M. Kawaguchi, Y. Takeuchi, An evolutionary hypothesis of binary opposition in functional incompatibility about habenular asymmetry in vertebrates. *Front. Neurosci.* **10**, 595 (2016).
69. B. C. Clarke, The evolution of genetic diversity. *Proc. R. Soc. Lond. B Biol. Sci.* **205**, 453–474 (1979).
70. D. R. Schield, B. W. Perry, R. H. Adams, M. L. Holding, Z. L. Nikolakis, S. S. Gopalan, C. F. Smith, J. M. Parker, J. M. Meik, M. DeGiorgio, S. P. Mackessy, T. A. Castoe, The roles of balancing selection and recombination in the evolution of rattlesnake venom. *Nat. Ecol. Evol.* **6**, 1367–1380 (2022).
71. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
72. C. Zhou, S. A. McCarthy, R. J. Durbin, YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
73. P. Kerpedjiev, N. Abdennur, F. Lekschas, C. McCallum, K. Dinkla, H. Strobelt, J. M. Luber, S. B. Ouellette, A. Azhir, N. Kumar, J. Hwang, S. Lee, B. H. Alver, H. Pfister, L. A. Mirny, P. J. Park, N. Gehlenborg, HiGlass: Web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
74. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, A. F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
75. Z. Bao, S. R. Eddy, Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276 (2002).
76. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acid Res.* **27**, 573–580 (1999).
77. S. Ou, N. Jiang, LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
78. D. Kuznetsov, F. Tegenfeldt, M. Manni, M. Seppey, M. Berkeley, E. V. Kriventseva, E. M. Zdobnov, OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acid Res.* **51**, D445–D451 (2023).
79. Y. Takeuchi, A. Ishikawa, Y. Oda, J. Kitano, Lateralized expression of left-right axis formation genes is shared by adult brains of lefty and righty scale-eating cichlids. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **28**, 99–106 (2018).
80. L. Gabriel, T. Brůna, K. J. Hoff, M. Ebel, A. Lomsadze, M. Borodovsky, M. J. G. R. Stanke, BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. bioRxiv 544449 [Preprint] (2024).
81. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
82. S. Kovaka, A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg, M. Pertea, Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
83. T. Brůna, A. Lomsadze, M. Borodovsky, GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**, 757–768 (2024).
84. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acid Res.* **34**, W435–W439 (2006).
85. L. Gabriel, K. J. Hoff, T. Brůna, M. Borodovsky, M. Stanke, TSEBRA: Transcript selector for BRAKER. *BMC Bioinformatics* **22**, 566 (2021).
86. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Mitchell, A. Quinn, A. Sangador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
87. C. P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
88. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
89. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
90. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
91. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
92. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
93. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
94. A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, R. Kikinis, 3D slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **30**, 1323–1341 (2012).
95. A. R. Palmer, C. Strobeck, FLUCTUATING ASYMMETRY: Measurement, analysis, patterns. *Annu. Rev. Ecol. Syst.* **17**, 391–421 (1986).
96. J. C. Gower, Generalized procrustes analysis. *Psychometrika* **40**, 33–51 (1975).
97. F. J. Rohlf, tpsDig, Digitize Landmarks and Outlines, version 2.05 (software) (2006).
98. K. Pearson, LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dubl. Phil. Mag. J. Sci.* **2**, 559–572 (1901).
99. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
100. D. Bu, H. Luo, P. Huo, Z. Wang, S. Zhang, Z. He, Y. Wu, L. Zhao, J. Liu, J. Guo, S. Fang, W. Cao, L. Yi, Y. Zhao, L. Kong, KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acid Res.* **49**, W317–W325 (2021).
101. M. Herrera, S. Vianello, L. Mitchell, Z. Chamot, C. Lorin-Nebel, E. Bonnelye, N. Roux, L. Besseau, Y. Gibert, V. Laudet, From genes to pathways: A curated gene approach to accurate pathway reconstruction in teleost fish transcriptomics. *J. Exp. Zool. B Mol. Dev. Evol.* 10.1002/jez.b.23299, (2025).

Acknowledgments: We thank J. Gerwin for help in Zambia in collecting fish in 2018, M. Pippel for assembling the genome as part of a broader collaboration between our laboratories, and G. Myers for providing funding for the genome assembly. We are grateful to the scientific computing cluster (SCC) and the animal research facility (TFA) of the University of Konstanz and S. Lee and J. Tuckermann from the University of Ulm for assistance with µCT. We thank R. Greenway and N. Roberto Ehemann for useful discussions. We thank N. Günther for help in the laboratory. **Funding:** This work was principally supported by DFG (German Science Foundation) to A.M. and a Hector Fellow Academy Ph.D. fellowship to X.T. **Author contributions:** X.T.: conceptualization, funding acquisition, formal analysis (data collection of genomic data and µCT scans, variant calling, balancing selection, genetic statistics, geometric morphometric analysis, and GWAS), visualization, and writing—original draft. M.L.: formal analysis (genome annotation) and writing—review and editing. A.M.: conceptualization, supervision, funding acquisition, and writing—review and editing. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the main text and/or the Supplementary Materials. The raw reads for the haplotype-resolved reference genome and all resequencing data from the *P. microlepis* have been deposited in the Sequence Read Archive (SRA; www.ncbi.nlm.nih.gov/sra) under project number PRJNA1218638. The two assembled haplotypes have been deposited in GenBank (www.ncbi.nlm.nih.gov/genbank/) with the accession numbers PRJNA1273515 and PRJNA1273516 and are also available via Zenodo (DOI: 10.5281/zenodo.15732779).

Submitted 5 February 2025
Accepted 30 June 2025
Published 30 July 2025
10.1126/sciadv.adw4406

Downloaded from https://www.science.org on July 31, 2025