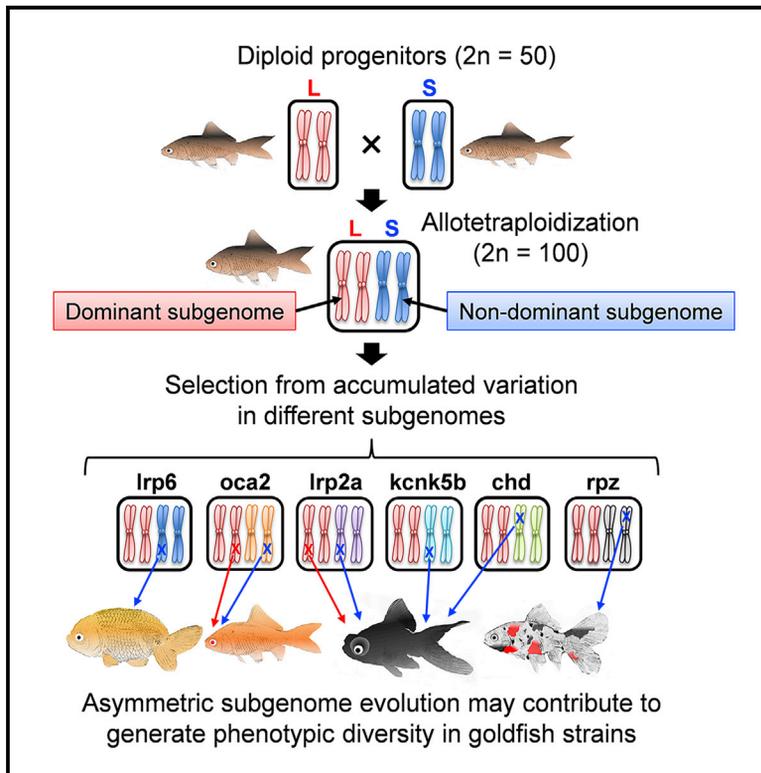# Current Biology

The Genetic Basis of Morphological Diversity in Domesticated Goldfish

Graphical Abstract

Authors

Tetsuo Kon, Yoshihiro Omori,
Kentaro Fukuta, ..., Shawn M. Burgess,
Hideki Noguchi, Takahisa Furukawa

Correspondence

yoshihiro.omori@protein.osaka-u.ac.jp

In Brief

Kon et al. provide genome-wide evidence of evolution and diversity of goldfish. Three groups were identified based on domestication history. Six genetic loci associated with five goldfish phenotypes are described, showing that differential evolution of subgenomes in goldfish is likely to have contributed to the diversity of goldfish phenotypes.

Highlights

- The goldfish genome was partitioned into two asymmetrically evolved subgenomes

- Wild goldfish and 27 goldfish strains were genetically classified into three groups

- Six genetic loci associated with five goldfish phenotypes were newly determined

- Asymmetrically evolved subgenomes seem to generate diverse goldfish phenotypes

CellPress

Current Biology

CellPress

Article

The Genetic Basis of Morphological Diversity in Domesticated Goldfish

Tetsuo Kon,[1] Yoshihiro Omori,[1,13,*] Kentaro Fukuta,[2] Hironori Wada,[3] Masakatsu Watanabe,[4] Zelin Chen,[5] Miki Iwasaki,[3] Tappei Mishina,[6,12] Shin-ichiro S. Matsuzaki,[7] Daiki Yoshihara,[1] Jumpei Arakawa,[8] Koichi Kawakami,[9] Atsushi Toyoda,[10] Shawn M. Burgess,[5] Hideki Noguchi,[2,11] and Takahisa Furukawa[1]

[1]Laboratory for Molecular and Developmental Biology, Institute for Protein Research, Osaka University, Osaka 565-0871, Japan
[2]Center for Genome Informatics, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Yata 1111, Mishima, Shizuoka 411-8540, Japan
[3]College of Liberal Arts and Sciences, Kitasato University, Sagamihara, Kanagawa, Japan
[4]Laboratory of Pattern Formation, Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka
[5]Translational and Functional Genomics Branch, National Human Genome Research Institute, Bethesda, MD, USA
[6]Laboratory of Animal Ecology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
[7]National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki, Japan
[8]Yatomi Station, Aichi Fisheries Research Institute, Yatomi, Aichi, Japan
[9]Laboratory of Molecular and Developmental Biology, National Institute of Genetics, and Department of Genetics, The Graduate University for Advanced Studies (SOKENDAI), Mishima, Japan
[10]Comparative Genomics Laboratory, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan
[11]Advanced Genomics Center, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan
[12]Present address: Laboratory for Chromosome Segregation, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe 650-0047, Japan
[13]Lead Contact
*Correspondence: yoshihiro.omori@protein.osaka-u.ac.jp
https://doi.org/10.1016/j.cub.2020.04.034

SUMMARY

Although domesticated goldfish strains exhibit highly diversified phenotypes in morphology, the genetic basis underlying these phenotypes is poorly understood. Here, based on analysis of transposable elements in the allotetraploid goldfish genome, we found that its two subgenomes have evolved asymmetrically since a whole-genome duplication event in the ancestor of goldfish and common carp. We conducted whole-genome sequencing of 27 domesticated goldfish strains and wild goldfish. We identified more than 60 million genetic variations and established a population genetic structure of major goldfish strains. Genome-wide association studies and analysis of strain-specific variants revealed genetic loci associated with several goldfish phenotypes, including dorsal fin loss, long-tail, telescope-eye, albinism, and heart-shaped tail. Our results suggest that accumulated mutations in the asymmetrically evolved subgenomes led to generation of diverse phenotypes in the goldfish domestication history. This study is a key resource for understanding the genetic basis of phenotypic diversity among goldfish strains.

INTRODUCTION

Goldfish (*Carassius auratus*), which belong to the family Cyprinidae, are a species that is closely related to the silver crucian carp; it was domesticated from wild goldfish during the Chinese Song dynasty (960–1279) [1]. Mitochondrial DNA analysis suggests that the domesticated goldfish was derived from a Chinese lineage of wild goldfish distinct from other *C. auratus* sublineages [2–4]. During the 1,000-year breeding history of goldfish strains, a wide variety of coloration and body, fin, eye, hood, and scale morphologies of the strains were established, mainly in East Asia. This wide variety of goldfish strains has long fascinated many researchers [1, 5, 6]. Charles Darwin was also interested in goldfish phenotypes and described the morphological features observed in goldfish strains [7]. Currently, at least 180 variants and 70 genetically established strains are produced and maintained all over the world [8]. Many goldfish strains display

characteristic phenotypes in hoods (epidermal thickening around the head), narial bouquets (hypertrophy of nasal septa), and caudal fins that have not been observed among mutant strains in other teleost models, including zebrafish and medaka [9], which indicates the unique mechanisms generating diverse phenotypes in goldfish. Furthermore, several goldfish strains display phenotypes similar to those of human diseases, including congenital glaucoma, skeletal abnormalities, and albinism [9]. Therefore, analyzing goldfish strains may lead to elucidation of the molecular mechanisms underlying vertebrate morphogenesis and certain human diseases.

Whole-genome duplication (WGD) is a doubling of the entire genome during the evolution of certain lineages and is hypothesized to provide a large amount of raw material for diversification and evolutionary innovations after generation of ohnologs (a pair of genes originating from WGD) [10]. Vertebrates are thought to have experienced two rounds (1R and 2R) of WGD

approximately 530–560 million years ago [11, 12]. In addition, a teleost-specific third-round WGD (Ts3R) occurred in a common ancestor of teleost fish approximately 320–350 million years ago [13–15]. The common ancestor of goldfish and common carp underwent a fourth round of WGD (Cs4R, carp-specific WGD) approximately 8–14 million years ago [9, 16–19]. Cs4R was suggested to be an allotetraploidization event, which is doubling of a complete set of chromosomes following interspecific hybridization of diploid progenitors (2n = 50) [20]. It has been reported, for certain allopolyploid plant species, that the two subgenomes have evolved asymmetrically after the WGD event [21, 22]. One of the two subgenomes is often preserved to stay more similar to the ancestral state, whereas the other experiences more chromosomal rearrangement, gene loss, and changes in levels of gene expression [23, 24]. Asymmetric subgenome evolution contributes to the phenotypic diversity of domesticated plants [21, 22, 25, 26]. In vertebrates, asymmetric subgenome evolution has also been observed in the genome of *Xenopus laevis* [23] and common carp (*Cyprinus carpio*) [27]. However, whether asymmetric subgenome evolution can contribute to genetic and phenotypic diversity in allopolyploid vertebrates remains unclear. Domesticated goldfish strains are an ideal model to clarify this issue because of their wide phenotypic diversity in morphology, revealing the phenotypic potential of an allopolyploid vertebrate under domestication.

Recently, we established a high-quality goldfish genome sequence using long-read sequencing technology [19]. Because of the relative evolutionary proximity of goldfish to the commonly used model organism zebrafish, comparative genome analysis is relatively easy between these species. It is also possible to examine goldfish gene functions in zebrafish models. This reference genome sequence provides a means to study the population genetic structure and molecular evolution of goldfish after Cs4R. In the current study, to clarify the relationship between subgenome evolution and phenotypic diversity of goldfish strains, we conducted whole-genome sequencing of goldfish strains displaying a wide variation in morphology and coloration. To the best of our knowledge, this is the first study to report the association between asymmetric subgenome evolution and phenotypic diversity in vertebrates.

RESULTS

Asymmetric Evolution of the Two Subgenomes

Cs4R has been suggested to be an allotetraploidization event based on cytological studies; elucidation of the subgenome structure is essential to determine whether Cs4R was an allotetraploidization or auto-tetraploidization event (doubling of a complete set of own chromosomes following polyploidization of a diploid progenitor species) [20]. In the previous study, 25 homeologous chromosome pairs were identified in the goldfish genome, but it was not partitioned into two homeologous subgenomes [19]. In the allotetraploid frog *X. laevis*, the transposable elements (TEs) specific to each progenitor have been reported to mark the descendent subgenomes [23]. Because the goldfish reference genome is assembled using long-read sequencing technology and is highly continuous [19], we adopted TE analysis to separate the goldfish reference genome into two homeologous subgenomes. We first defined the subgenome-biased

index (SBI) as follows (Figure S1A). For each homeologous chromosome pair, the absolute value of the difference in the number of a given TE between two homeologous chromosomes was calculated. Then the sum of these values was divided by the total number of the given TE. The SBI takes a value from 0 to 1, and a value close to 1 indicates that the TE is expected to show subgenome-biased distribution. We calculated SBI values for 2,114 types of TEs and identified 22 types of TEs with subgenome-biased distributions (Figures S1B and S1C; Data S1A). According to the distribution of these 22 types of TEs on the 50 goldfish chromosomes, we partitioned the goldfish chromosomes into two subgenomes: L (relatively long) and S (relatively short) (Figures 1A, S1D, and S1E). The L chromosomes were, on average, 5.2 ± 3.4 Mb longer than the S chromosomes in the reference sequence (p = 0.003916, paired two-sided Student's t test). Twenty types of TEs, including hAT-N91 and Tc1-1, were almost exclusively located in chromosomes of the L subgenome (Figures 1A and S1D). In contrast, the other two types of TEs, Mariner-12 and Mariner-N17, were almost exclusively located in chromosomes of the S subgenome (Figures 1A and S1D). Our phylogenetic analysis revealed that these 22 types of TEs exhibited extensive sequence divergence, suggesting that these TEs are relicts of ancient TEs (Figures S1F–S1L). We calculated the ratio of the nonsynonymous substitution rate per nonsynonymous site (dN) to the synonymous substitution rate per synonymous site (dS) between zebrafish orthologs and goldfish ohnolog pairs (Figures 1A, S2A, and S2B). The median dN/dS in the S subgenome was significantly larger than that in the L subgenome. We also noted that gene loss and gene disruption were more common in the S subgenome than in the L subgenome (Figures S2C–S2F). Furthermore, genes located in the L subgenome tended to have higher expression levels in brain, eye, heart, muscle, bone, gill, and tail fin tissues than those located in the S subgenome (Figure S2G). These results suggest that Cs4R was an allotetraploidization event, that the L and S subgenomes evolved asymmetrically, and that the L subgenome is the dominant subgenome, which is more often preserved more similar to the ancestral state [26, 28].

Recently, it has been reported that the common carp has two subgenomes: subgenomes A and B [27]. Because the common ancestor of goldfish and common carp underwent Cs4R, we compared the genomes of these two species. The protein coding sequences of 5,404 ohnolog pairs in goldfish and those of 8,291 ohnolog pairs in common carp were compared, and reciprocal best hits were identified. We found that most genes on the L chromosomes in goldfish were exclusively mapped to the B chromosomes in common carp (Figure S2H). Similarly, most genes on the S chromosomes in goldfish were exclusively mapped to the A chromosomes in common carp. A goldfish L subgenome marker, hAT-N91, showed highly biased distributions to B chromosomes in the common carp genome (Figure S2I, top panel). Similarly, a goldfish S subgenome marker, Mariner-17, is located biasedly in the A subgenome in the common carp genome (Figure S2I, bottom panel). We found that singleton genes were observed more frequently in the L subgenome than in the S subgenome (Figure S2C). Similarly, it has been reported that, in common carp, singleton genes were observed more frequently in subgenome B, which is orthologous to the L subgenome in goldfish [27]. To observe the differences in
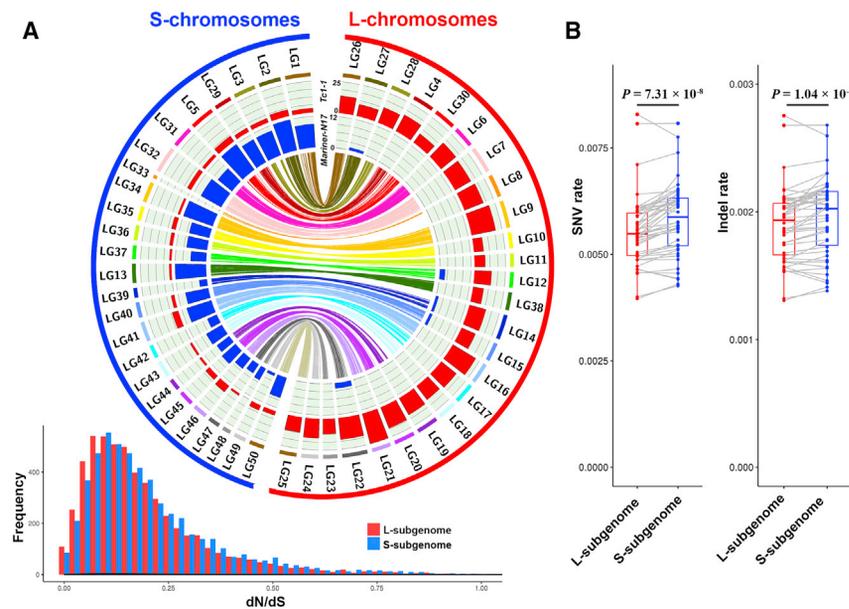
Current Biology
Article

CellPress

Figure 1. Asymmetric Subgenome Evolution of Goldfish

(A) Top panel: division of the goldfish genome into two subgenomes based on the biased distribution of TEs. The goldfish genome was divided into two subgenomes, the L subgenome (right half) and the S subgenome (left half), based on the biased distributions of TEs such as *Tc1-1* (outer track, red) and *Mariner-N17* (inner track, blue). *Tc1-1* is almost exclusively located in the L subgenome (right half). *Mariner-N17* is almost exclusively located in the S subgenome (left half). The links represent 5,404 pairs of goldfish ohnologs. Bottom panel: the dN/dS distribution of ohnologs in the L subgenome (red) and the S subgenome (blue). We focused on 5,404 gene pairs located in goldfish LG1–LG50 and zebrafish chromosomes 1–25. Each aligned ohnolog gene was compared independently with its zebrafish ortholog, and the ratio of nonsynonymous substitution rate per nonsynonymous site (dN) to synonymous substitution rate per synonymous site (dS) was calculated. The greater the difference in dN/dS, the more selection was relaxed on one gene in the pair. The median of dN/dS of the S subgenome was significantly larger than that of the L subgenome (p = 1.48 × 10^{-40}, paired Wilcoxon rank-sum test).

(B) SNV (left) and indel (right) frequencies in the L or S subgenome among the 27 goldfish strains. Both frequencies were significantly higher in the S subgenome (blue) than in the L subgenome (red) (paired two-sided Student's t test).

See also Figures S1 and S2 and Data S1.

ohnolog evolution after Cs4R between goldfish and common carp, we investigated whether the common carp orthologs of goldfish singletons exist as singletons or ohnologs and vice versa. We found that 41% of goldfish singletons also exist as singletons in the common carp genome and that 59% of goldfish singleton genes have two or more copies in the common carp genome. Conversely, 64% of common carp singletons exist as singletons in the goldfish genome, and 36% of common carp singleton genes have two or more copies in the goldfish genome.

In the current study, we named duplicated genes after Cs4R "gene L" or "gene S" according to the localized subgenome; in addition, we distinguished genes that were duplicated after Ts3R as "gene a" and "gene b." For example, we named the four paralogs of *kcnk5* in the goldfish genome *kcnk5aL*, *kcnk5aS*, *kcnk5bL*, and *kcnk5bS*.

Whole-Genome Sequencing of Goldfish Strains

To clarify the genetic diversity, population structure, and phylogenetic relationships among goldfish strains and wild goldfish, we chose strains displaying a large variety of phenotypes in morphology, including body coloration (Figures S3A–S3Y; Data S1B and S1C), and conducted whole-genome sequencing of 48 individuals from 27 strains bred in Japan and a wild goldfish lineage (obtained from the Lake Kasumigaura and Lake Hinuma basins, C6 clade in Wang et al. [4]). We obtained 637 Gb of sequence (Data S1D) and aligned sequence reads against the goldfish reference genome [19]. We identified 63,347,761 variants across the 48 goldfish individuals (Data S1D and S1E). To clarify whether the asymmetric genetic diversity between the two subgenomes is observed among the goldfish strains, we investigated the single-nucleotide variant (SNV)

and insertion or deletion (indel) frequency of each subgenome in the dataset containing 27 goldfish strains. Both frequencies were significantly higher in the S than in the L subgenome (0.553% in the L subgenome and 0.581% in the S subgenome for SNVs; 0.189% in the L subgenome and 0.195% in the S subgenome for indels) (Figure 1B), showing that there is asymmetric genetic diversity between the two subgenomes in domesticated goldfish strains.

Using these data, we analyzed the population structure of the goldfish strains (Figures 2A–2D, S3Z, and S4A–S4E). In the admixture analysis, we found that the cross-validation error was lowest at the number of ancestries (K) = 3 (red, green, and blue; Figures 2B and S3Z). Next we classified the 48 goldfish individuals into three groups according to the proportion of the individual's genome from inferred ancestral populations at K = 3. We named these groups to reflect the history of goldfish breeding: China (blue), Ranchu (green), and Edo (red) (Figures 2A, 2B, and S3Z'; Data S1C). Consistent with the admixture analysis, principal-component analysis (PCA) also separated the individuals into three groups (Figures S4A–S4C). Furthermore, the maximum-likelihood tree and the neighbor-joining tree of 48 individuals also supported this grouping (Figures 2C, S4D, and S4E). To search the regions of reduced heterogeneity and potential fixation in different lineages, we calculated fixation index (Fst) values in 40-kb windows sliding 10 kb at a time for the China, Ranchu, and Edo groups. We identified 1,069 regions (total 90 Mb), 1,128 regions (total 110 Mbp), and 2,096 regions (total 236 Mb) with high Fst values for the China, Ranchu, and Edo groups (Data S1F). For example, the regions from 7.1–8.3 Mb on LG31, 12.1–12.2 Mb on LG15, and 15.3–16.8 Mb on LG19 showed high Fst values, respectively (Figures S4F–S4H). These
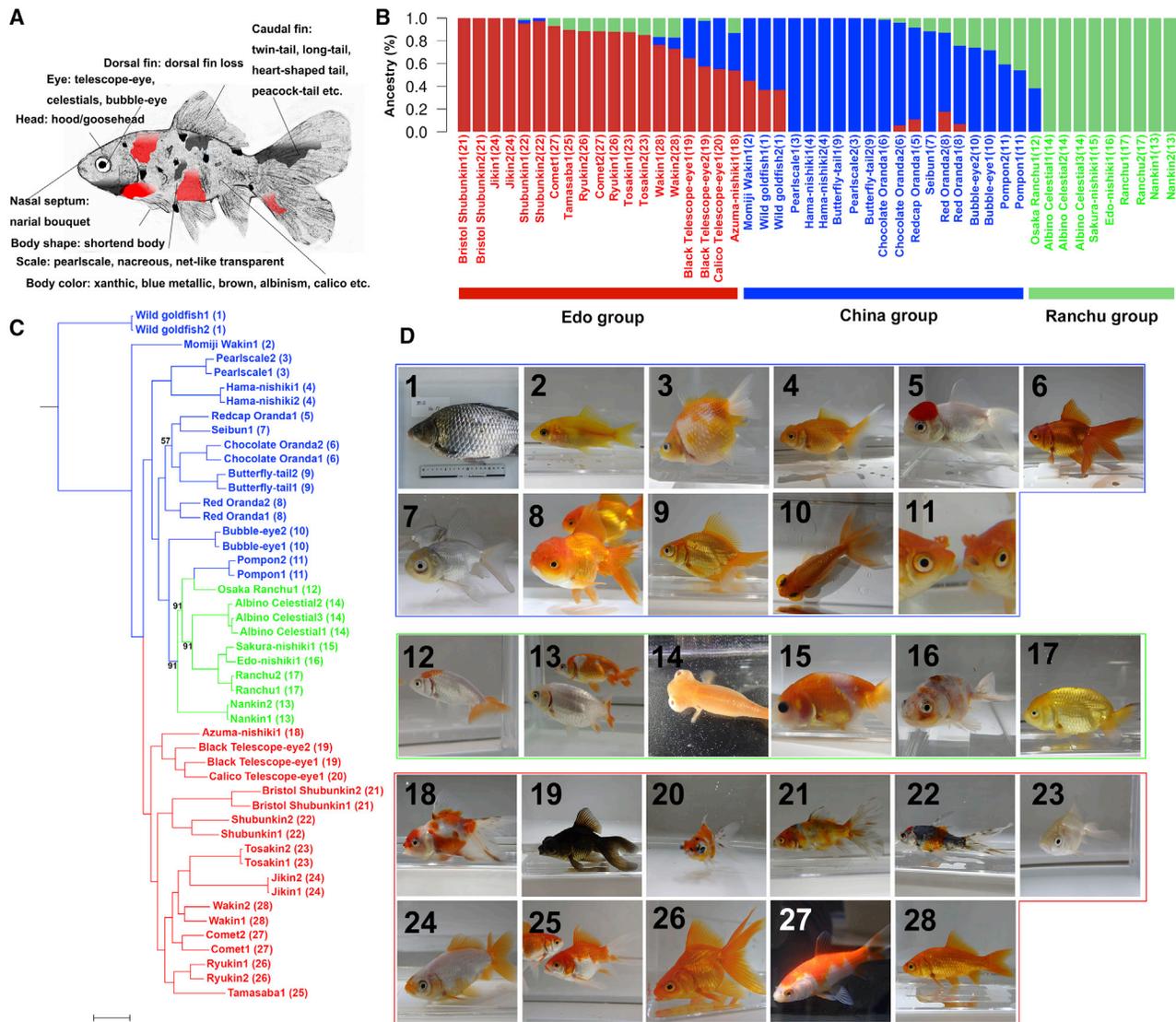
Figure 2. The Genetic Population Structure of 27 Goldfish Strains and Wild Goldfish

(A) Overview of the goldfish phenotypes in the 27 goldfish strains analyzed in this study. The details of the phenotypes of the goldfish strains are presented in Figures S3A–S3Q and Data S1B. The drawing shows a lateral view of a Bristol Shubunkin goldfish.

(B) Admixture analysis of the 48 goldfish individuals (K = 3). The x axis indicates the goldfish individuals, and the y axis indicates estimated ancestry proportions. Based on this analysis, the 48 individuals were classified into three groups: Edo (n = 19, red), China (n = 19, blue), and Ranchu (n = 10, green).

(C) Maximum-likelihood tree of the analyzed goldfish individuals. The scale bar represents 0.05 substitutions per base pair. Only bootstrap values less than 95% are shown.

(D) Images of the 27 goldfish strains and the wild goldfish analyzed in this study. The names of each strain and their associated numbers are shown in (B) and (C). See also Figures S3 and S4 and Data S1.

regions are considered to be areas of reduced heterogeneity and potential fixation among strains in each group. It is possible that these include regions of reduced heterogeneity because of a domestication bottleneck that was also observed in other domesticated animals [29].

The genomic regions with a high degree of fixation in goldfish strains may include genes that were positively selected during the history of goldfish domestication [30, 31]. To identify such loci, we calculated the Z-transformed pooled heterogeneity (ZH_p) in 40-kb windows sliding 10 kb at a time (Figures S4I and

S4J). In this analysis, we identified reduced heterogeneity regions containing 2,020 genes (Figure S4K; Data S1G). The Gene Ontology (GO) analysis identified significantly enriched GO terms for this gene set (Data S1H). Notably, the most enriched GO term was "negative regulation of signal transduction" (GO: 0009968, 2.36-fold enrichment, false discovery rate [FDR] value = 0.00248). We also identified the highly enriched GO term "embryonic organ development" (GO: 0048568, 1.95-fold enrichment, FDR value = 0.00829) for this gene group. These enrichments of GO terms in lower ZH_p regions may have
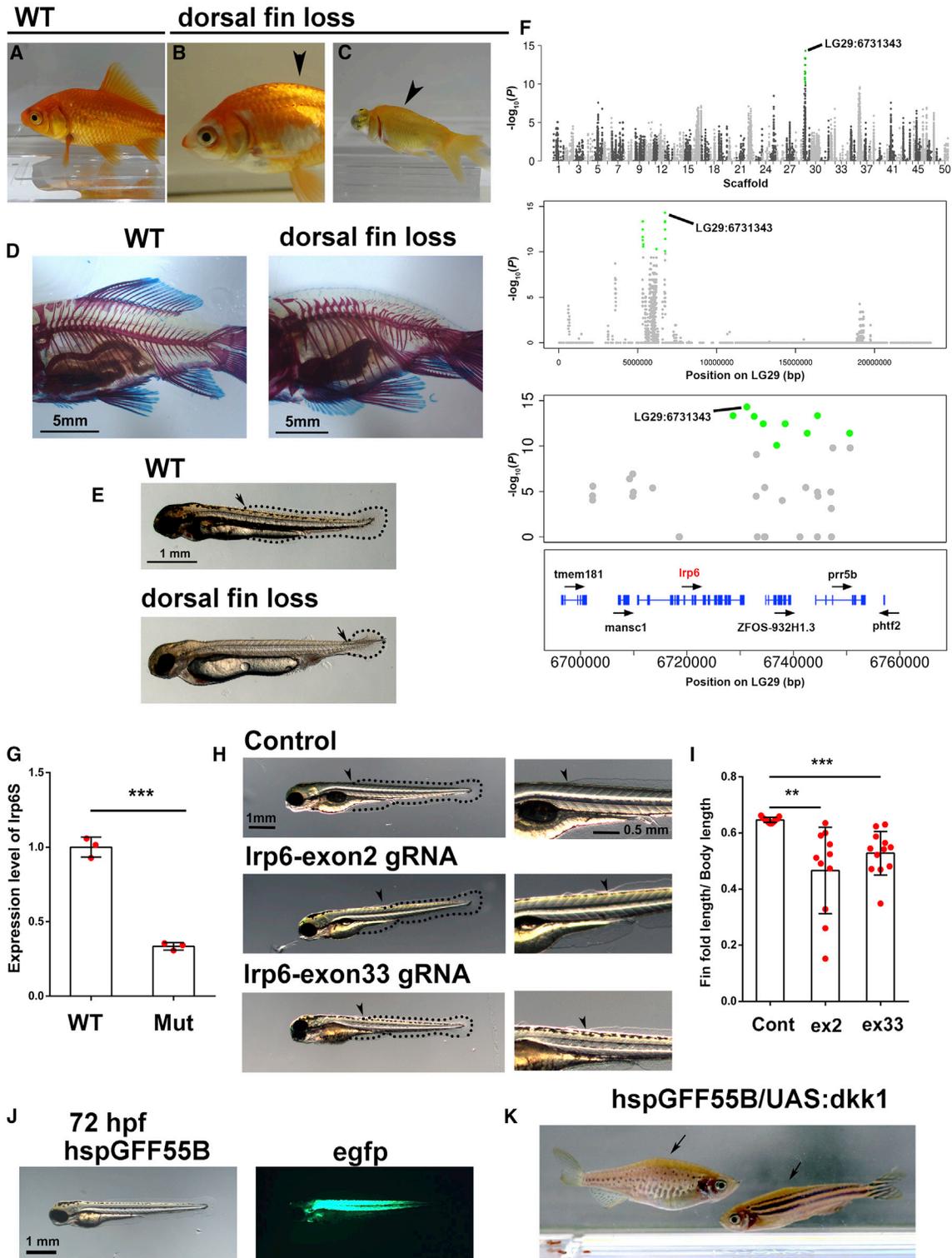
CellPress

Figure 3. Association of the Dorsal Fin Loss Phenotype with the _Irp6S_ Locus on LG29

(A) A normal dorsal fin observed in common goldfish.

(B and C) The dorsal fin loss phenotype observed in Osaka Ranchu (B) and Celestial (C) goldfish.

(D) Alcian blue and alizarin red skeletal staining of common goldfish (left panel) and Albino Ranchu (right panel). The Albino Ranchu goldfish showed complete ablation of the fin rays and endoskeletal elements of the dorsal fin (common goldfish, n = 10; Albino Ranchu, n = 10).

(E) In common goldfish, larvae develop the extended fin fold at 3 dpf stage (top panel). In Albino Ranchu embryos, most of the dorsal fin fold was lost (bottom panel) (common goldfish, n = 30; Albino Ranchu, n = 30). The start position of the dorsal fin fold is indicated by arrowheads.

undergone purifying selection for developmental genes because of the genes' key function in development.

Genome-wide Association Study of Goldfish Phenotypes

To identify genetic loci associated with phenotypes observed in goldfish strains, we performed genome-wide association study (GWAS) on seven representative goldfish phenotypes using the variant genomic datasets identified in the current study (Data S1I). The results revealed loci that were highly associated with the twin-tail, dorsal fin loss, long-tail, and telescope-eye phenotypes.

The Twin-Tail Phenotype

Many goldfish strains, including Ryukin, Oranda, and Telescope-eye, possess a bifurcated caudal axial skeleton, presenting as the twin-tail phenotype (Figures 2D; Data S1B). A mutation in one of the *chordin* paralogs (*chdS*) has been reported to cause the twin-tail phenotype in goldfish [32]. However, whether the other *chordin* paralog (*chdL*), an ohnolog, contributes to the twin-tail phenotype remains unknown. To test whether our GWAS detected mutations in *chdL* and/or *chdS*, we conducted GWAS using the data of 21 twin-tail strains as cases and seven single-tail strains and wild goldfish as controls. Our GWAS of the twin-tail phenotypes revealed the highest association between the genomic variants on LG40 at base positions 19,923,396 and 19,923,744 and the twin-tail phenotype (p = 6.583 × 10^{-15}) (Figures S5A–S5C). Combined with a synteny analysis using zebrafish and medaka genome sequences, we identified the *chdS* gene in this region. We observed no significant association between the locus *chdL* on LG15 and the twin-tail phenotype, indicating that *chdL* does not significantly contribute to the twin-tail phenotype, at least in the strains tested. This analysis showed that the mutation in *chordin* of twin-tail goldfish is on the non-dominant S subgenome and that whole-genome sequencing and GWAS analysis based on the subgenome structures are effective for identifying loci responsible for phenotypes in goldfish strains.

The Dorsal Fin Loss Phenotype

Several goldfish strains, including the Ranchu group, exhibit a dorsal fin loss phenotype (Figures 2D and 3A–3C). The mode of inheritance of this phenotype is most likely recessive [33]. To investigate the detailed skeletal structure of the dorsal fin, we performed Alcian blue and alizarin red staining of common goldfish with a normal dorsal fin and of Albino Ranchu goldfish, which displays the dorsal fin loss phenotype. Common goldfish (single-tailed Red Wakin strain) retain the body form and fins of the wild

goldfish; however, it has red coloration. We used common goldfish as the wild type for comparing phenotypes with other strains. We observed complete loss of dorsal fin rays and endoskeletal elements (distal and proximal radials) in Albino Ranchu goldfish (Figure 3D). We also examined the dorsal fin fold formation at the larval stage and found that most of the dorsal fin fold in Albino Ranchu goldfish was lost 3 days post fertilization (dpf) (Figure 3E).

To identify the loci associated with the dorsal fin loss phenotype, we conducted GWAS using eight strains, with the dorsal fin loss phenotype as cases and the other 19 strains and wild goldfish as controls. Our GWAS identified the highest association between a genomic variant at LG29 position 6,731,343 and the dorsal fin loss phenotype (p = 4.893e × 10^{-15}) (Figure 3F). A 26-kb homozygous haplotype (6,728,745–6,755,075) containing three genes in all eight goldfish strains with the dorsal fin loss phenotype was identified around this position (Data S1J). Within this 26-kb region, we identified an Wnt co-receptor Lrp6S as a strong candidate responsible for the dorsal fin loss phenotype because Wnt signaling is essential for fin formation as well as fin regeneration in teleost fish [34–36]. It has also been reported that partial knockdown of *lrp6* in *Xenopus* embryos affects formation of the fin fold at the larval stage [37]. We searched for mutations in *lrp6S* of goldfish with the dorsal fin loss phenotype and identified four amino acid substitutions; however, all of them were conserved or semi-conserved in zebrafish or medaka *lrp6* orthologs (Figure S5D), suggesting that these *lrp6S* missense mutations were unlikely to be causative mutations of the dorsal fin loss phenotype. The fin fold of goldfish embryos starts forming at the 25% otic vesicle closure (OVC) stage [38]. We analyzed the expression level of *lrp6S* at the 25% OVC stage and discovered that *lrp6S* expression notably decreased in Albino Ranchu embryos (Figure 3G), suggesting that a mutation in the regulatory elements of *lrp6S* may cause the dorsal fin loss phenotype in goldfish. We identified a 313-bp deletion in intron 21 of *lrp6S* in goldfish with the dorsal fin loss phenotype (6,729,459–6,729,771) (Figure S5E). This intronic deletion might affect induction and proper production of *lrp6S* mRNA at embryonic stages.

Because a goldfish genome editing method has not yet been established, it is technically difficult to observe the *lrp6* loss-of-function phenotype in goldfish. We used zebrafish, which belong to the carp family, to observe the effect of loss of function of *lrp6*. To investigate whether partial loss of *lrp6* function affects fin formation in teleost fish, we performed CRISPR-Cas9-mediated gene disruption of *lrp6* in zebrafish. We identified a partial defect of dorsal fin fold formation in 4 dpf zebrafish larvae injected with

---

(F) GWAS identified the association of the dorsal fin loss phenotype with the *lrp6S* locus on LG29.

(G) The *lrp6S* expression level in Albino Ranchu goldfish embryos at the 25% OVC stage significantly decreased compared with common goldfish (p = 8.4 × 10^{-5}, n = 3, two-sided Student's t test). Each circle represents a pool of embryos.

(H and I) Partial loss of the dorsal portion of the fin fold was observed in zebrafish larvae injected with *lrp6* gRNA and Cas9 (H). The start position of the dorsal fin fold is indicated by arrowheads. The ratio of the unaffected dorsal fin fold length to body length significantly decreased in 4 dpf zebrafish larvae injected with *lrp6* gRNA and Cas9 (I). Control versus exon 2 mutations, p = 0.0030, two-sided Welch's t test; control versus exon 33 mutations, p = 0.00025, two-sided Welch's t test. Each circle represents an individual or a pool of embryos. Control, n = 13; exon 2 mutations, n = 11; exon 33 mutations, n = 12. Cont, control; ex, exon.

(J and K) Partial inhibition of the Lrp6-mediated Wnt pathway by ectopic expression of Dkk1 affects dorsal fin formation at the adult stage. Shown is expression of EGFP in muscular tissues of hspGFF55B/UAS:EGFP larvae at 72 hpf (J). Adult hspGFF55B/UAS:dkk1a-rfp zebrafish lost the dorsal fin (K), similar to Ranchu goldfish.

***p < 0.001, **p < 0.01, *p < 0.05; mean ± SD.

See also Figure S5 and Data S1.

Current Biology
Article

CellPress

lrp6 guide RNA (gRNA) and Cas9 (Figures 3H and S5F). The ratio of unaffected dorsal fin fold length to body length was significantly reduced in these embryos (Figure 3I). *Lrp6* knockout in mice or knockdown in zebrafish or *Xenopus* causes severe embryonic developmental defects [39–41]. The remaining *lrp6S* expression and/or expression of the ohnolog *lrp6L* seem to prevent lethality in goldfish with the dorsal fin loss phenotype. Mosaic biallelic gene disruption by CRISPR-Cas9-mediated genome editing occurs in F0 zebrafish larvae [42]. Mosaic gene disruption in zebrafish larvae is most likely to prevent lethality, although knockdown of *lrp6* causes severe embryonic defects in zebrafish [41].

To analyze whether partial inhibition of the Wnt pathway mediated by Lrp6 affects dorsal fin formation in adult teleost fish, we generated zebrafish that ectopically expressed the Wnt inhibitor Dkk1 after the dorsal fin fold formation stage. Dkk1 negatively regulates Lrp6-mediated Wnt signaling via direct interaction with the Wnt co-receptor Lrp6 in vertebrates [43–45]. We observed that zebrafish larvae of the Gal4 driver line expressed enhanced green fluorescent protein (EGFP) in the dorsal part of the trunk 72 h post fertilization (hpf) (Figures 3J, S5G, and S5H). We found that larvae expressing *dkk1* under control of an upstream activating sequence (UAS) lost the dorsal part of the fin fold (Figures S5G and S5H), which was similar to the larval phenotype of Albino Ranchu goldfish (Figure 3E). Notably, at the adult stage, we found that the dorsal fin was lost in these lines (Figure 3K). These results suggest that the Lrp6-mediated Wnt pathway regulates dorsal fin formation in adult zebrafish. These results support the hypothesis that the causative mutation for the phenotype of dorsal fin loss is around the *lrp6S* gene locus on the S subgenome and that decreased *lrp6S* expression affects adult dorsal fin formation in goldfish.

The Long-Tail Phenotype

Goldfish strains with the long-tail phenotype exhibit notably elongated caudal fin lobes in single- and twin-tailed goldfish strains (Figures 4A–4E). Goldfish with the long-tail phenotype usually show elongation of all median and paired fins, suggesting that this mutation modulates the length of all fin types. The long-tail phenotype in goldfish is most likely to be dominant [33]. A GWAS was conducted using a dataset of 14 strains with the long-tail phenotype as cases and 13 strains without the long-tail phenotype and wild goldfish as controls. This analysis identified a notable association between a locus on LG45 at base position 15,015,973 and the long-tail phenotype (p = 3.358 × 10^{-15}) (Figures S5I–S5K). In the vicinity of this position, all 14 tested strains with the long-tail phenotype carried at least one copy of a 46-kb long haplotype spanning four genes (14,942,902–14,988,949; Data S1K). In this region, we identified *kcnk5bS* as a strong candidate gene because mutations in its zebrafish ortholog (another longfin [*alf*]) have been reported to cause a proportionally larger caudal fin phenotype similar to the goldfish long-tail phenotype [46]. KCNK5 is a member of the two-pore domain potassium channel family that produces background (leak) K$^+$ currents over a large range of membrane potentials [47, 48]. *alf* mutations increase the K$^+$ conductance of the channel to cause hyperpolarization of cells, although it is unknown how the change in K$^+$ channel conductance regulates fin ray length [46]. Our sequence analysis identified five amino acid substitutions or deletions in

kcnk5bS in goldfish with the long-tail phenotype (Figure S6A). Interestingly, one of these mutations, V165E, was in the vicinity of the mutated amino acid W169L, reported in the zebrafish *alf* mutant (Figure 4F). To test whether mutations in goldfish *kcnk5bS* cause a K$^+$ conductance change, we performed voltage-clamp recordings using *Xenopus* oocytes injected with cRNA of wild-type and mutant goldfish *kcnk5bS*. K$^+$ conductance in oocytes injected with mutant *kcnk5bS* cRNA significantly increased compared with oocytes injected with wild-type *kcnk5bS* cRNA (Figures 4G and 4H), suggesting that the goldfish mutant *kcnk5bS* causes significant cell hyperpolarization. Homology modeling of the kcnk5bS structure revealed that the V165E mutation in the goldfish kcnk5bS was localized in the M3 transmembrane domain (Figures 4I, S6B, and S6C). Substitution of a hydrophobic amino acid (valine) with a hydrophilic amino acid (glutamic acid) possibly caused a critical alteration of channel gating, supporting the hypothesis that the *kcnk5bS* mutation on the S subgenome is a causal mutation of the goldfish long fin phenotype.

The Telescope-Eye Phenotype

Certain goldfish strains possess enlarged protuberant eyes, known as the telescope-eye phenotype (Figures 5A–5D). A previous genetic study of Telescope-eye goldfish identified this phenotype as recessive [33]. However, the gene locus or mutations causing the telescope-eye phenotype are unknown. To identify the locus associated with the telescope-eye phenotype in goldfish, we conducted a GWAS of goldfish strains using datasets of four Telescope-eye strains as cases and the other 23 strains and wild goldfish with normal eyes as controls. Our GWAS analysis revealed a mild association (p = 1.29 × 10^{-7}) between a genomic variant on LG9 at base position 5,399,083 and the telescope-eye phenotype (Figures S6D and S6E). In a region relatively close to this locus, we identified *lrp2aL* as a candidate gene for the telescope-eye phenotype because *Lrp2a* mutations have been shown to cause an enlarged eye phenotype in adult zebrafish and mice that is similar to the goldfish telescope-eye phenotype [49, 50]. In mouse models, LRP2 plays the role of sonic hedgehog (shh) clearance receptor and regulates shh-induced cell proliferation at the retinal margin, resulting in the large eye phenotype [50]. A detailed sequence analysis of *lrp2aL* variants in strains with the telescope-eye phenotype identified two nonsense mutations in *lrp2aL* genes that were only present in the genome of goldfish with the telescope-eye phenotype and absent in other goldfish (Figures 5E and S6F). We also performed RNA sequencing (RNA-seq) analysis to compare the gene expression profile in the eyes of the common goldfish with that of the Black Telescope-eye goldfish. Intriguingly, we detected a notable *lrp2aL* transcript abnormality in the Black Telescope-eye strain, where the *lrp2aL* mRNA lacked half of its 3′ portion (Figures S6G and S6H). The breakpoint was located between exon 45 and exon 46. Analysis of this region revealed a 13-k base pairs (bp) insertion in intron 45 in the Black Telescope-eye strain, but this was absent in common goldfish. Nucleotide sequencing analysis revealed that the insertion encoded a TE that is a type of foamy-like endogenous retrovirus (Figures 5E and S6I–S6K). The *de novo* transcriptome assembly of the RNA-seq data identified that an aberrant *lrp2aL* mRNA expressed in the Black Telescope-eye strain possessed a premature stop codon (Figures S6G and S6H). Almost no normal *lrp2aL* mRNA was expressed
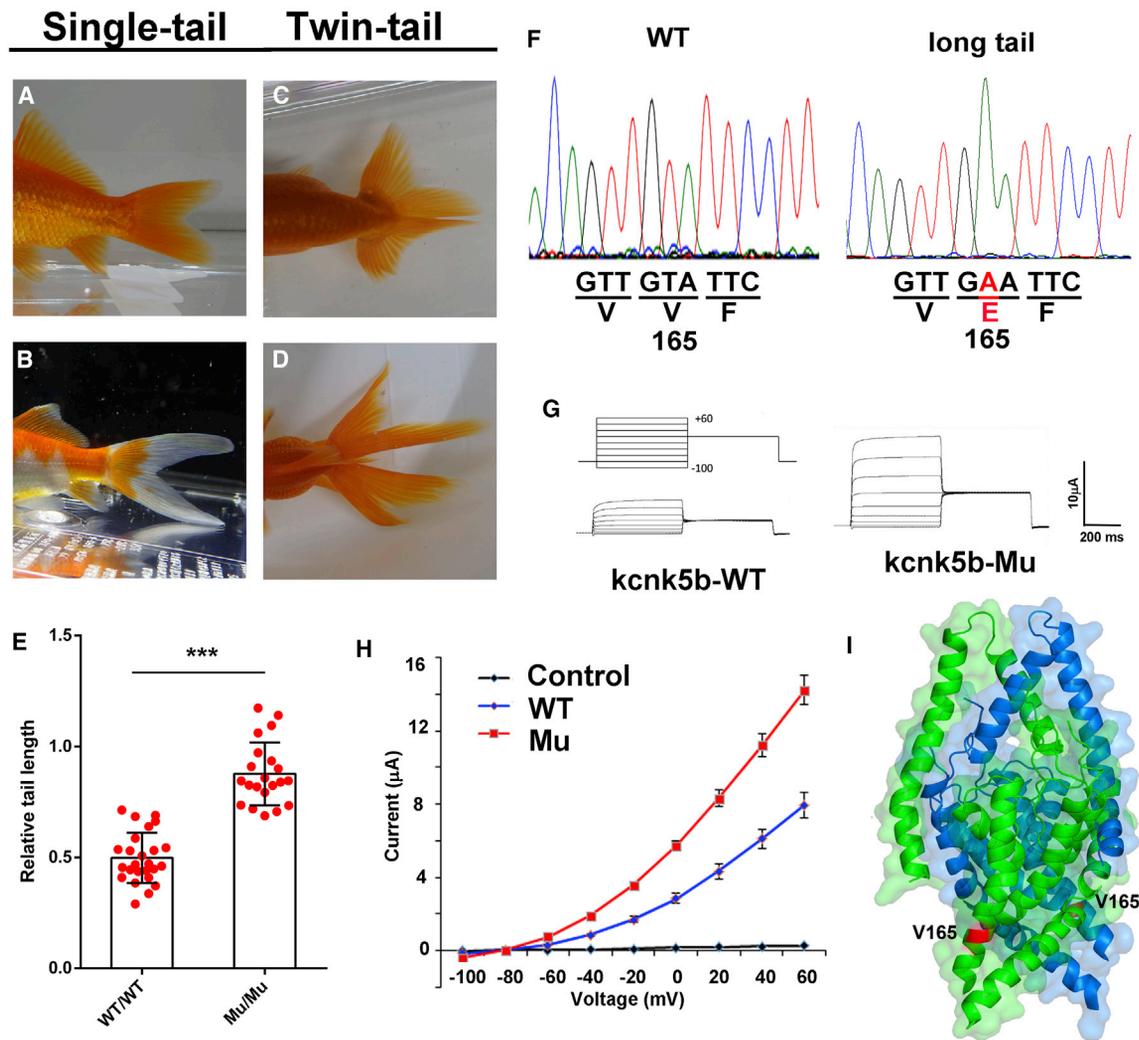
Figure 4. Mutations in *kcnk5bS* Are Linked to the Long Fin Phenotypes

(A–E) The long-tail phenotype (B and D) is observed in the single tail (A and B) and twin-tail (C and D) strains. The caudal fin length of the goldfish strains with the long-tail phenotype was significantly longer compared with that of goldfish strains with a normal fin (E) ($p = 2.8 \times 10^{-13}$; wild-type, n = 25; mutant, n = 21; two-sided Student's t test). ***$p < 0.001$, * mean ± SD. WT, wild-type; Mu, mutant.

(F) A missense mutation (V165E) was identified in goldfish with the long-tail phenotype.

(G) Representative electrophysiological traces of voltage-clamp recordings using *Xenopus* oocytes injected with cRNA (complementary RNA) of goldfish WT and Mu *kcnk5b*.

(H) The K$^+$ conductance in oocytes injected with Mu cRNA increased significantly compared with oocytes injected with WT cRNA.

(I) The predicted 3D structure of goldfish Kcnk5b. The position of the mutated amino acid (V165) is indicated in red.

See also Figures S5 and Figure S6 and Data S1.

in the Black Telescope-eye strain (<0.1% of that found in common goldfish), suggesting that the 13-kbp retrotransposon insertion was responsible for the loss of *lrp2aL* function in Telescope-eye goldfish. Overall, we identified three types of loss-of-function mutations in the *lrp2aL* gene in strains with the telescope-eye phenotype. These results support the hypothesis that loss-of-function mutation in *lrp2aL* on the L subgenome causes the telescope-eye phenotype in goldfish.

Analysis of Strain-Specific Variants

Several goldfish phenotypes observed in this study were only present in a single strain among the 27 we analyzed (Data

S1B), such as albinism, which was only observed in the Albino Celestial strain. Thus, it is possible that such phenotypes are caused by strain-specific mutations. To explore this further, we searched for strain-specific homozygous variants (SSVs) in the 18 goldfish strains that had a strain-specific phenotype, identifying 7,023 genes with SSVs (Figure S7A; Data S1L and S1M; Table S1).

Albinism

Many goldfish strains with xanthic bodies lack melanophores in the skin and scales but retain black retinal pigment epithelia in the retina. This is because the pigment cells in the body surface
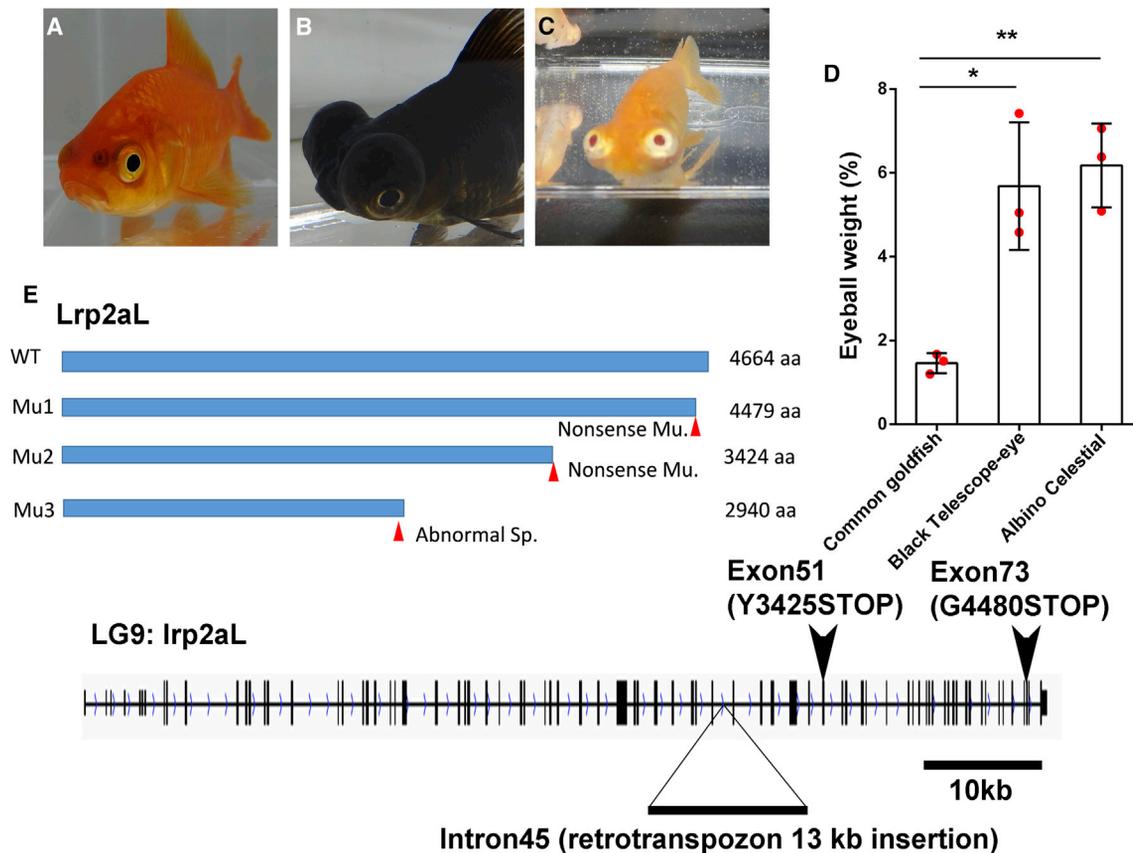
CellPress

Figure 5. Mutations in *lrp2aL* Are Linked to the Telescope-Eye Phenotypes

(A–C) Enlarged and protuberant eyeballs are observed in the Black Telescope-eye (B) and Albino Celestial (C) strains compared with normal eyeballs of common goldfish (A). The eyes of Albino Celestial goldfish are upwardly directed (C).

(D and E) The eye weight of Black Telescope-eye (Bl) and Albino Celestial goldfish (Ce) was increased significantly compared with common goldfish (Co) (D). Each circle represents an individual (Co versus Bl, p = 0.038; Co, n = 3; Bl, n = 3; two-sided Welch's t test and Co versus Ce, p = 0.0014; Co, n = 3; Ce, n = 3; two-sided Student's t test). **p < 0.01, *p < 0.05; mean ± SD Three types of mutations were identified in goldfish strains with the telescope-eye phenotype. (E) The 13-kb retrotransposon insertion was identified in intron 45 of *lrp2aL* in strains with the telescope-eye phenotype.

See also Figure S6 and Data S1.

are derived from neural crest cells, whereas retinal pigment epithelia originate from optic lobe neuroepithelial cells [51]. These goldfish have black pupils that reflect the black retinal pigment epithelia. However, in albino goldfish, the melanin loss occurs on the body surface and in the retina, resulting in pink pupils because of depigmentation of the retinal pigment epithelia (Figures 6A, 6B, S7B, and S7C) and loss of dark body coloration in juveniles.

Although a previous genetic study reported that albinism in goldfish is double recessive for two independently assorting autosomal loci, *m/m* and *s/s* [52], the mutated genes are unknown. To identify mutations associated with the albino phenotype, we searched the 172 genes with SSVs in Albino Celestial goldfish (Data S1L) for any overlap with the 143 genes reported as being related to body color in vertebrates (Data S1N) and found that only *oca2* overlapped between these two groups. Interestingly, the critical mutations identified in both *oca2* ohnologs (*oca2L* and *oca2S*) of Albino Celestial goldfish had frameshift indels, resulting in production of truncated proteins comprising 470 and 519 amino acids, respectively (Figures 6C, S7D, and S7E). We observed that none of the 26 non-albino

strains possessed the frameshift mutation in *oca2L* or the homozygous frameshift mutation in *oca2S*. We also analyzed mutations of *oca2* ohnologs in the Albino Telescope-eye, Albino Azuma-nishiki, Albino Comet, Albino Ranchu, and Albino Oranda. We found that all of these strains possessed the same homozygous mutations in both *oca2* ohnologs, similar to Albino Celestial goldfish. Because *oca2* encodes an anion transporter regulating the pH of the melanosome, and because loss-of-function mutations of *oca2* in several species, including humans, mice, and zebrafish, are known to lead to albinism [53–55], these mutations in *oca2* ohnologs on the L and S subgenomes are most likely to be the causative mutations of albinism in goldfish strains.

The Heart-Shaped Tail Phenotype

The heart-shaped caudal fin phenotype was only displayed in Bristol Shubunkin goldfish among the 27 strains analyzed in this study. To clarify the heart-shaped tail phenotype, we analyzed the fin ray length of the caudal fins (Figure S7F). Adult common goldfish have 10 and 9 principal caudal fin rays in upper and lower lobes, respectively [56] (Figure S7F). We measured the
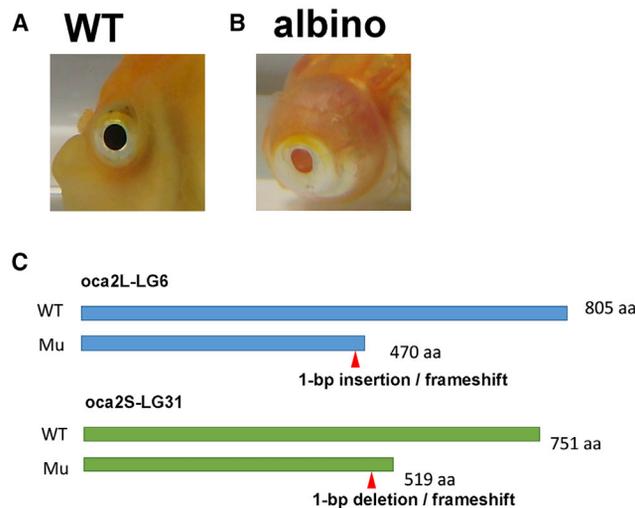
Figure 6. Identification of the Mutations in the Albino Strains
(A) The normal black pupil of Ranchu goldfish.
(B) The Albino Telescope-eye has pink pupils.
(C) The frameshift mutations found in the *oca2* ohnologs in the albino strains. See also Figure S7, Table S1, and Data S1.

length of the tenth caudal fin ray in the dorsal side of the caudal fin cleft (10 dfr), the second caudal fin ray in the dorsal fin lobe (2 dfr), and the sixth dorsal caudal fin ray (6 dfr). In most cases, 2 dfr was the longest ray in the caudal fin of the dorsal lobe. We observed that the ratio of 2 dfr length/body length in Shubunkin and Bristol Shubunkin goldfish was significantly larger compared with common goldfish (Figure S7G) (p = 0.0029, control n = 6; Shubunkin n = 4, two-sided Welch's t test; p = 1.9×10^{-6}, common goldfish n = 6, Bristol Shubunkin n = 4, two-sided Student's t test). Notably, the ratio of 6 dfr/2 dfr in Bristol Shubunkin goldfish was significantly higher than that of Shubunkin goldfish (p = 0.0077; Shubunkin, n = 4; Bristol Shubunkin, n = 4; two-sided Student's t test), suggesting that regulation of the caudal fin ray length proportion was affected in Bristol Shubunkin goldfish.

We identified 1,466 genes with SSVs for Bristol Shubunkin goldfish (Data S1L). To identify the candidate genes responsible for the heart-shaped tail phenotype, we searched for overlapping genes between genes reported previously to regulate fin shape and genes with SSVs in the Bristol Shubunkin strain. We found that the Bristol Shubunkin SSVs contained the *rpz* gene cluster. In zebrafish, a teleost-specific transmembrane protein, Rpz, regulates caudal fin length [57], although its molecular mechanism is still unelucidated. The zebrafish *rpz* gene cluster is reported to contain five *rpz*-like genes [57]. Our detailed analysis revealed seven zebrafish *rpz*-like genes in this cluster. In goldfish, we found two *rpz* gene clusters on LG16 (L subgenome) and LG41 (S subgenome). Each of them contains seven *rpz*-like genes similar to those in zebrafish. We observed that *rpzS* and *rpz4S* were relatively highly expressed in the caudal fin (Data S1O). We found a Bristol Shubunkin-specific missense mutation (p.Asp124Glu) in *rpzS* and two (Tyr133Phe and Phe319Ser) in *rpz4S*. Further analysis is needed to clarify whether the *rpz* gene cluster in the S subgenome contributes to the heart-shaped tail phenotype. Interestingly, a previous proteomics analysis reported asymmetric distribution of Rpz and Rpz5 in zebrafish caudal fins; these

proteins are significantly enriched in the proximal region of the caudal fin in zebrafish [51]. A calcineurin-mediated mechanism is hypothesized to operate as a molecular switch between the position-associated isometric and allometric growth program in caudal fins [58]. Thus, the asymmetric distribution of Rpz-like proteins may contribute to fin shape formation by regulating the balance between the isometric and allometric growth programs. Our results support the hypothesis that mutations in the *rpz* gene cluster on the S subgenome affect the fin ray length proportion in caudal fins and contribute to the heart-shaped tail phenotype in Bristol Shubunkin goldfish.

DISCUSSION

Why do goldfish exhibit so much diversity in their morphology and body colors? Our results provide a clue to answering this question. Previous studies suggested that gene duplications may facilitate the emergence of new adaptive functions of the duplicated genes because the original function can be retained in other duplicated copies [10, 59, 60]. Therefore, WGD and excess gene duplications in the genome may profoundly affect species diversification. Similarly, duplicated genes in the two subgenomes possibly contribute to generation of diverse goldfish phenotypes during selection by breeders. In the current study, we identified five new loci associated with goldfish phenotypes (long-tail, telescope-eye, dorsal fin loss, albinism, and heart-shaped tail). Combined with the reported locus (twin-tail), our study has identified the chromosomal loci linked to six phenotypes in goldfish (Figure 7). Our TE analysis partitioned the goldfish chromosomes into L and S subgenomes. In our study, telescope-eye is the only phenotype associated with only the L subgenome locus. Both *oca2* ohnologs on the L and S subgenomes were mutated in strains with the albino phenotype. The other four phenotypes were associated with loci only in the S subgenome. In the S subgenome, we identified more gene loss, gene disruption, and sequence divergence than in the L subgenome (Figures 1A and S2A–S2F). Notably, we observed that the frequencies of SNVs and indels were higher in the S than in the L subgenome among the 27 goldfish strains (Figure 1B). These results suggest that the two goldfish subgenomes evolved asymmetrically, similar to those in *X. laevis* [23], and that there is a wider variation of SNVs and indels in the S than in the L subgenome. Because there are fewer established phenotypes of *X. laevis* or common carp than of goldfish, goldfish could be an excellent model to clarify phenotype-genotype relationships in animals that underwent WGD. We propose a model in which diversified mutations accumulated in ohnologs in the asymmetrically evolved subgenomes frequently generate diverse morphological and body color phenotypes that are frequently selected by goldfish breeders (Figure 7). If neofunctionalization and/or subfunction loss of ohnologs is more frequent in the S than in the L subgenome, the novel morphological phenotypes of goldfish caused by mutations in the S subgenome may be more frequent than those in the L subgenome. Therefore, the existence of two subgenomes with asymmetric evolution in the goldfish genome possibly contributes to production of a higher number of varieties in morphology and body color phenotypes than in other teleost fish. In plants, a similar phenomenon has been reported in a domesticated allotetraploid cotton that has
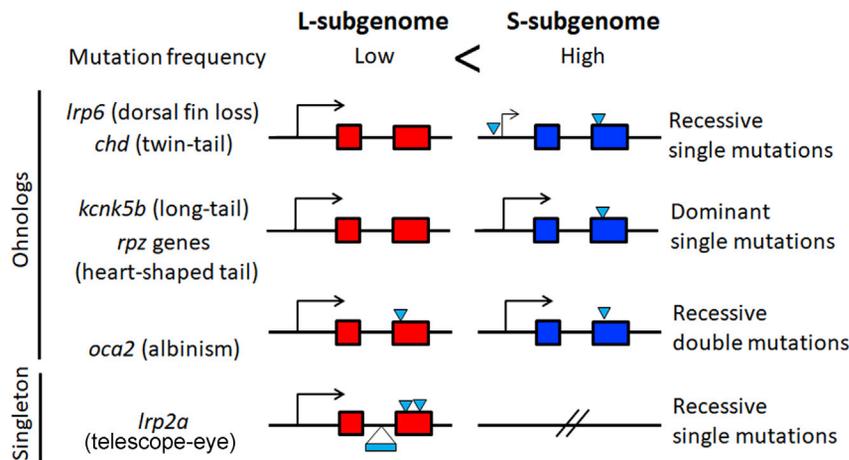
CellPress

Figure 7. A Diagram Showing the Asymmetric Subgenome Evolution and the Candidate Genes for the Goldfish Phenotypes in the Two Subgenomes

The modes of inheritance of the phenotypes are indicated on the right. See details in Discussion. See also Figure S7.

of these species could still be improved. Future work in which both genomes are annotated with the same pipeline and parameters will further elucidate the divergent evolution of goldfish and common carp after Cs4R.

In the present study, we first report whole-genome sequencing of goldfish strains with a variety of morphological and body color phenotypes. We identify more than 60 million genetic variations, including SNVs and small indels, across 48 goldfish individuals. Based on these genetic variants in the nuclear genome of the strains, show, for the first time, a population structure of goldfish strains (Figure 2). These results classify the major goldfish strains into three groups. This study reveals critical information about the basic genetic background of major goldfish strains with diverse phenotypes and is a key resource for further analysis of the genetic basis of phenotypic diversity among goldfish strains. Information about the genes with SSVs may help link phenotype to genotype in goldfish strains with specific morphology and/or coloration in future studies.

All domesticated goldfish used in this study were bred and developed in Japanese traditional breeding regions, including Yatomi, Hamamatsu, Saitama, and Nagano. Ancestral strains of these goldfish were imported from China to Japan at various times between the Muromachi and Showa periods (1336–1989) [5, 33]. The ancestral strains were modified by crossing different strains, and the current strains were established (described in detail in Data S1C and Figure S3Z'). Previous population genetic analyses of goldfish strains have been reported based on mitochondrial genome variations [3, 4]. In this study, we analyzed variations in the nuclear genome and revealed the detailed population structure of goldfish strains. Our analysis segregated the major goldfish strains into three groups, China, Ranchu, and Edo, which reflect the geographical history of goldfish breeding [1, 5, 33]. For example, between the Muromachi and Meiji periods (1336–1912), the ancestral strains of the Edo group (Wakin, Ryukin, Jikin, Tosakin, Tamasaba, Comet, Telescope-eye, Shubunkin, and Bristol Shubunkin) were imported from China to Japan and bred and modified in Japan [5, 33]. We found that 15 individuals of strains of the Edo group, with the exceptions of Shubunkin and Bristol Shubunkin, are grouped in the PCA plot of the first two components of the 48 goldfish individuals (Figure S4B), reflecting the long history of crossing within these strains and/or the existence of common ancestral strains in Japan. There is a record that the common ancestor of the Shubunkin and Bristol Shubunkin strains was generated from Calico Telescope-eye goldfish, common goldfish, and Japanese crucian carp around 1900 in Japan [33]. In line with this scenario, we observed that individuals of the Shubunkin and Bristol

the A subgenome and the D subgenome, described as having undergone asymmetric subgenome evolution, although the mechanism underlying this is unclear [61]. There are more genes positively selected during cotton domestication in the A subgenome (470 genes) than in the D subgenome (341 genes). Further analysis is required to clarify whether such asymmetric subgenome evolution during domestication occurred in allotetraploid plants and animals via common or different mechanisms.

Among the candidate genes for the six phenotypes that we studied here, only *kcnk5* and *lrp2* have paralogs after the Ts4R event. In the goldfish genome, we found four paralogs of *kcnk5* (*kcnk5aL*, *kcnk5aS*, *kcnk5bL*, and *kcnk5bS*). The *lrp2* genes in goldfish are most likely lost paralogs on the S subgenome after Cs4R (Figure S7H), and only two paralogs were identified (*lrp2aL* and *lrp2bL*). We compared the amino acid sequences of candidate gene paralogs (Data S1P). We observed that the paralogs S and L show clearly higher identities (80%–92% amino acid identity) than those between paralogs a and b (44%–56% amino acid identity), reflecting the ancient history of Cs4R and Ts3R [11–15], consistent with previous studies [62–65]. We also compared the expression of these paralogs (Data S1O). We observed that ohnologs of *lrp6*, *kcnk5a*, and *rpz* on the S subgenome are expressed more than those on the L subgenome in the tissues examined. In contrast, *chdL* is expressed more than *chdS* in these tissues. The dominant expression of these ohnologs seems to have been established after Cs4R.

The two diploid progenitors of common carp and goldfish diverged 23 million years ago (mya) and merged 12.4 mya (Cs4R) [27]. By comparing the goldfish genome with the common carp genome, we found that all L and S chromosomes in goldfish were orthologous to chromosomes of subgenomes B and A in common carp, suggesting that the outline of the subgenome structure is retained between the goldfish and common carp genomes. We also observed that singletons in the goldfish genome do not partially overlap with those in the common carp genome. This partially non-overlapping composition of singletons between the common carp and goldfish genomes may reflect divergent evolution of these two species after Cs4R. Because we used the goldfish and common carp genome annotations generated based on different annotation pipelines, it should be noted that the comparability of genome annotations

Shubunkin strains are mapped apart from all other domesticated goldfish individuals in the PCA plot of the 48 goldfish individuals, suggesting a history of crossing with Japanese crucian carp (Figure S4A).

We performed a GWAS for the seven phenotypes (Data S1I). Among them, we could not obtain a significant peak of GWAS for the phenotypes, including short body, hood, and calico. It is possible that these phenotypes are caused by multiple causative genes and/or multiple alleles in the goldfish strains tested. In future studies, analyses by whole-genome sequencing of more individuals and/or strains may reveal loci for these phenotypes. Alternatively, genome sequencing analyses of the hybrid strains generated by crossing between different strains could contribute to determining the causative loci for these phenotypes.

Several goldfish genes associated with the phenotypes identified in this study have been reported to have human orthologs with mutations in genetic disorders. In humans, mutations resulting in *LRP2* loss cause Donnai-Barrow syndrome [66]. *OCA2* mutations lead to oculocutaneous albinism type II [67, 68]. Thus, goldfish strains with mutations observed in human genetic disorders are possibly useful model animals for these diseases.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
- METHOD DETAILS
  - Partitioning the goldfish genome into two homeologous subgenomes
  - Phylogenetic analysis of the subgenome-specific TEs
  - Molecular evolutionary analysis of goldfish subgenomes
  - Comparative genome analysis between goldfish and common carp
  - Purification of genomic DNA and total RNA from the goldfish tissues
  - Whole-genome sequencing and RNA-seq of the goldfish strains and wild goldfish
  - Read trimming and mapping
  - Variant calling, filtering, and annotations
  - Estimation of the subgenome-specific SNV and indel rates between the goldfish strain genomes and the reference genome
  - Admixture analysis and grouping of the goldfish strains
  - Principal component analysis (PCA)
  - Phylogenetic analysis of the goldfish strains and the wild goldfish
  - Fst analysis
  - Pooled heterogeneity of the goldfish strains
  - Identification of strain-specific variants
  - Phylogenetic analysis of the goldfish strains, wild goldfish, crucian carp, and common carp
  - Genome-wide association study (GWAS)
  - Homology modeling of the goldfish Kcnk5b structure

- RNA-seq analysis
- Identification of gFV-1 transposable element insertion in lrp2aL of Black Telescope-eye goldfish
- Phylogenetic analysis of gFV-1_NA, gFV, and their derivatives
- Synteny analysis of the lrp2aL locus in LG9 and LG34
- Candidate gene list for pigmentation and fin shape
- Skeletal staining of goldfish
- CRISPR/Cas9-mediated target mutagenesis in zebrafish
- Partial inhibition of Wnt signaling by ectopic expression of Dkk1
- Voltage clamp recording
- Quantitative RT-PCR (qRT-PCR) analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.04.034.

AUTHOR CONTRIBUTIONS

T.K., A.T., and Y.O. designed the project. T.K., K.F., H.W., M.W., Z.C., M.I., T.M., S.S.M., D.Y., J.A., K.K., T.F., H.N., S.M.B., A.T., and Y.O. performed experiments and analyzed data. T.K., K.F., H.W., M.W., A.T., T.F., and Y.O. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Chen, S. (1956). A history of the domestication and the factors of the varietal formation of the common goldfish, Carassius auratus. Sci. Sin. 5, 287–321.

2. Takase, Y., Fujitani, H., and Murakami, M. (2008). Isolation and Characterization of Genomic DNA Markers from the Gynogenetic Triploid Ginbuna (Carassius auratus langsdrofi) by Amplified Fragment Length Polymorphism Analysis. Journal of Azabu University 15, 1–7.

Current Biology
Article

CellPress

3. Komiyama, T., Kobayashi, H., Tateno, Y., Inoko, H., Gojobori, T., and Ikeo, K. (2009). An evolutionary origin and selection process of goldfish. Gene 430, 5–11.

4. Wang, S.Y., Luo, J., Murphy, R.W., Wu, S.F., Zhu, C.L., Gao, Y., and Zhang, Y.P. (2013). Origin of Chinese goldfish and sequential loss of genetic diversity accompanies new breeds. PLoS ONE 8, e59571.

5. Smartt, J. (2008). Goldfish Varieties and Genetics: Handbook for Breeders (Blackwell Science).

6. Kirpicnikov, V.S. (1987). Genetische Grundlagen der Fischzüchtung (VEB Deutscher Landwirtschaftsverlag).

7. Darwin, C. (1868). The variation of animals and plants under domestication Volume 2 (Murray).

8. Nasu, M., and Ohuchi, Y. (2016). Nishikigoi and Goldfish (Seibundo Shinkosha).

9. Omori, Y., and Kon, T. (2019). Goldfish: an old and new model system to study vertebrate development, evolution and human disease. J. Biochem. 165, 209–218.

10. Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. 10, 725–732.

11. Dehal, P., and Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3, e314.

12. Kasahara, M. (2007). The 2R hypothesis: an update. Curr. Opin. Immunol. 19, 547–552.

13. Alsop, D., and Vijayan, M. (2009). The zebrafish stress axis: molecular fallout from the teleost-specific genome duplication event. Gen. Comp. Endocrinol. 161, 62–66.

14. Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711–1714.

15. Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. 13, 382–390.

16. Larhammar, D., and Risinger, C. (1994). Molecular genetic aspects of tetraploidy in the common carp Cyprinus carpio. Mol. Phylogenet. Evol. 3, 59–68.

17. David, L., Blum, S., Feldman, M.W., Lavi, U., and Hillel, J. (2003). Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci. Mol. Biol. Evol. 20, 1425–1434.

18. Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., et al. (2014). Genome sequence and genetic diversity of the common carp, Cyprinus carpio. Nat. Genet. 46, 1212–1219.

19. Chen, Z., Omori, Y., Koren, S., Shirokiya, T., Kuroda, T., Miyamoto, A., Wada, H., Fujiyama, A., Toyoda, A., Zhang, S., et al. (2019). De Novo assembly of the goldfish (Carassius auratus) genome and the evolution of genes after whole genome duplication. Sci. Adv. 5, eaav0547.

20. Ohno, S., Muramoto, J., Christian, L., and Atkin, N.B. (1967). Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae. Chromosoma 23, 1–9.

21. Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. USA 108, 4069–4074.

22. Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., Ma, J., Yu, J., Huang, S., et al. (2014). The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. 5, 3930.

23. Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., et al. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538, 336–343.

24. Mirzaghaderi, G., and Mason, A.S. (2017). Revisiting Pivotal-Differential Genome Evolution in Wheat. Trends Plant Sci. 22, 674–684.

25. Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., et al. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nat. Genet. 49, 579–587.

26. Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. Nat. Plants 4, 258–268.

27. Xu, P., Xu, J., Liu, G., Chen, L., Zhou, Z., Peng, W., Jiang, Y., Zhao, Z., Jia, Z., Sun, Y., et al. (2019). The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio. Nat. Commun. 10, 4625.

28. Hu, G., and Wendel, J.F. (2019). Cis-trans controls and regulatory novelty accompanying allopolyploidization. New Phytol. 221, 1691–1700.

29. Wang, G.D., Xie, H.B., Peng, M.S., Irwin, D., and Zhang, Y.P. (2014). Domestication genomics: evidence from animals. Annu. Rev. Anim. Biosci. 2, 65–84.

30. Axelsson, E., Ratnakumar, A., Arendt, M.L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, A., and Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495, 360–364.

31. Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464, 587–591.

32. Abe, G., Lee, S.H., Chang, M., Liu, S.C., Tsai, H.Y., and Ota, K.G. (2014). The origin of the bifurcated axial skeletal system in the twin-tail goldfish. Nat. Commun. 5, 3360.

33. Matsui, Y. (1935). Kagaku to Shumi Kara Mita Kingyo no Kenkyuu (Seizando Syoten).

34. Kawakami, Y., Rodriguez Esteban, C., Raya, M., Kawakami, H., Martí, M., Dubova, I., and Izpisúa Belmonte, J.C. (2006). Wnt/beta-catenin signaling regulates vertebrate limb regeneration. Genes Dev. 20, 3232–3237.

35. Nagayoshi, S., Hayashi, E., Abe, G., Osato, N., Asakawa, K., Urasaki, A., Horikawa, K., Ikeo, K., Takeda, H., and Kawakami, K. (2008). Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: tcf7 and synembryn-like. Development 135, 159–169.

36. Tatsumi, Y., Takeda, M., Matsuda, M., Suzuki, T., and Yokoi, H. (2014). TALEN-mediated mutagenesis in zebrafish reveals a role for r-spondin 2 in fin ray and vertebral development. FEBS Lett. 588, 4543–4550.

37. Hassler, C., Cruciat, C.M., Huang, Y.L., Kuriyama, S., Mayor, R., and Niehrs, C. (2007). Kremen is required for neural crest induction in Xenopus and promotes LRP6-mediated Wnt signaling. Development 134, 4255–4263.

38. Tsai, H.Y., Chang, M., Liu, S.C., Abe, G., and Ota, K.G. (2013). Embryonic development of goldfish (Carassius auratus): a model for the study of evolutionary change in developmental mechanisms by artificial selection. Dev. Dyn. 242, 1262–1283.

39. Mathew, L.K., Sengupta, S.S., Ladu, J., Andreasen, E.A., and Tanguay, R.L. (2008). Crosstalk between AHR and Wnt signaling through R-Spondin1 impairs tissue regeneration in zebrafish. FASEB J. 22, 3087–3096.

40. Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J., and Skarnes, W.C. (2000). An LDL-receptor-related protein mediates Wnt signalling in mice. Nature 407, 535–538.

41. Jiang, Y., He, X., and Howe, P.H. (2012). Disabled-2 (Dab2) inhibits Wnt/β-catenin signalling by binding LRP6 and promoting its internalization through clathrin. EMBO J. 31, 2336–2349.

42. Jao, L.E., Wente, S.R., and Chen, W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. Proc. Natl. Acad. Sci. USA 110, 13904–13909.

43. Hashimoto, H., Itoh, M., Yamanaka, Y., Yamashita, S., Shimizu, T., Solnica-Krezel, L., Hibi, M., and Hirano, T. (2000). Zebrafish Dkk1 functions in forebrain specification and axial mesendoderm formation. Dev. Biol. 217, 138–152.

44. Mao, B., Wu, W., Li, Y., Hoppe, D., Stannek, P., Glinka, A., and Niehrs, C. (2001). LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. Nature *411*, 321–325.

45. Zebisch, M., Jackson, V.A., Zhao, Y., and Jones, E.Y. (2016). Structure of the Dual-Mode Wnt Regulator Kremen1 and Insight into Ternary Complex Formation with LRP6 and Dickkopf. Structure *24*, 1599–1605.

46. Perathoner, S., Daane, J.M., Henrion, U., Seebohm, G., Higdon, C.W., Johnson, S.L., Nüsslein-Volhard, C., and Harris, M.P. (2014). Bioelectric signaling regulates size in zebrafish fins. PLoS Genet. *10*, e1004080.

47. Enyedi, P., and Czirják, G. (2010). Molecular background of leak K+ currents: two-pore domain potassium channels. Physiol. Rev. *90*, 559–605.

48. Lesage, F., and Barhanin, J. (2011). Molecular physiology of pH-sensitive background K(2P) channels. Physiology (Bethesda) *26*, 424–437.

49. Veth, K.N., Willer, J.R., Collery, R.F., Gray, M.P., Willer, G.B., Wagner, D.S., Mullins, M.C., Udvadia, A.J., Smith, R.S., John, S.W., et al. (2011). Mutations in zebrafish lrp2 result in adult-onset ocular pathogenesis that models myopia and other risk factors for glaucoma. PLoS Genet. *7*, e1001310.

50. Christ, A., Christa, A., Klippert, J., Eule, J.C., Bachmann, S., Wallace, V.A., Hammes, A., and Willnow, T.E. (2015). LRP2 Acts as SHH Clearance Receptor to Protect the Retinal Margin from Mitogenic Stimuli. Dev. Cell *35*, 36–48.

51. Bharti, K., Nguyen, M.T., Skuntz, S., Bertuzzi, S., and Arnheiter, H. (2006). The other pigment cell: specification and development of the pigmented epithelium of the vertebrate eye. Pigment Cell Res. *19*, 380–394.

52. Yamamoto, T.-o. (1973). Inheritance of albinism in the goldfish, Carassius auratus. Jpn. J. Genet. *48*, 53–64.

53. Lee, S.T., Nicholls, R.D., Schnur, R.E., Guida, L.C., Lu-Kuo, J., Spinner, N.B., Zackai, E.H., and Spritz, R.A. (1994). Diverse mutations of the P gene among African-Americans with type II (tyrosinase-positive) oculocutaneous albinism (OCA2). Hum. Mol. Genet. *3*, 2047–2051.

54. Brilliant, M.H. (2001). The mouse p (pink-eyed dilution) and human P genes, oculocutaneous albinism type 2 (OCA2), and melanosomal pH. Pigment Cell Res. *14*, 86–93.

55. Beirl, A.J., Linbo, T.H., Cobb, M.J., and Cooper, C.D. (2014). oca2 Regulation of chromatophore differentiation and number is cell type specific in zebrafish. Pigment Cell Melanoma Res. *27*, 178–189.

56. Li, I.J., Chang, C.J., Liu, S.C., Abe, G., and Ota, K.G. (2015). Postembryonic staging of wild-type goldfish, with brief reference to skeletal systems. Dev. Dyn. *244*, 1485–1518.

57. Green, J., Taylor, J.J., Hindes, A., Johnson, S.L., and Goldsmith, M.I. (2009). A gain of function mutation causing skeletal overgrowth in the rapunzel mutant. Dev. Biol. *334*, 224–234.

58. Kujawski, S., Lin, W., Kitte, F., Börmel, M., Fuchs, S., Arulmozhivarman, G., Vogt, S., Theil, D., Zhang, Y., and Antos, C.L. (2014). Calcineurin regulates coordinated outgrowth of zebrafish regenerating fins. Dev. Cell *28*, 573–587.

59. Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature *513*, 375–381.

60. Paape, T., Briskine, R.V., Halstead-Nussloch, G., Lischer, H.E.L., Shimizu-Inatsugi, R., Hatakeyama, M., Tanaka, K., Nishiyama, T., Sabirov, R., Sese, J., and Shimizu, K.K. (2018). Patterns of polymorphism and selection in the subgenomes of the allopolyploid Arabidopsis kamchatica. Nat. Commun. *9*, 3909.

61. Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., et al. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. *33*, 531–537.

62. Omori, Y., and Malicki, J. (2006). oko meduzy and related crumbs genes are determinants of apical cell features in the vertebrate embryo. Curr. Biol. *16*, 945–957.

63. Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., et al. (2016). Corrigendum: The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. *48*, 700.

64. Abe, G., Lee, S.H., Li, I.J., and Ota, K.G. (2018). An alternative evolutionary pathway for the twin-tail goldfish via szl gene mutation. J. Exp. Zoolog. B Mol. Dev. Evol. *330*, 234–241.

65. Caetano-Lopes, J., Henke, K., Urso, K., Duryea, J., Charles, J.F., Warman, M.L., and Harris, M.P. (2020). Unique and non-redundant function of csf1r paralogues in regulation and evolution of post-embryonic development of the zebrafish. Development *147*, dev.181834.

66. Kantarci, S., Al-Gazali, L., Hill, R.S., Donnai, D., Black, G.C., Bieth, E., Chassaing, N., Lacombe, D., Devriendt, K., Teebi, A., et al. (2007). Mutations in LRP2, which encodes the multiligand receptor megalin, cause Donnai-Barrow and facio-oculo-acoustico-renal syndromes. Nat. Genet. *39*, 957–959.

67. Durham-Pierre, D., Gardner, J.M., Nakatsu, Y., King, R.A., Francke, U., Ching, A., Aquaron, R., del Marmol, V., and Brilliant, M.H. (1994). African origin of an intragenic deletion of the human P gene in tyrosinase positive oculocutaneous albinism. Nat. Genet. *7*, 176–179.

68. Lee, S.T., Nicholls, R.D., Bundey, S., Laxova, R., Musarella, M., and Spritz, R.A. (1994). Mutations of the P gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism. N. Engl. J. Med. *330*, 529–534.

69. Ruboyianes, R., and Worobey, M. (2016). Foamy-like endogenous retroviruses are extensive and abundant in teleosts. Virus Evol. *2*, vew032.

70. Koga, A., and Hori, H. (2001). The Tol2 transposable element of the medaka fish: an active DNA-based element naturally occurring in a vertebrate genome. Genes Genet. Syst. *76*, 1–8.

71. Jiang, X.Y., Du, X.D., Tian, Y.M., Shen, R.J., Sun, C.F., and Zou, S.M. (2012). Goldfish transposase Tgf2 presumably from recent horizontal transfer is active. FASEB J. *26*, 2743–2752.

72. Dong, Y.Y., Pike, A.C., Mackenzie, A., McClenaghan, C., Aryal, P., Dong, L., Quigley, A., Grieben, M., Goubin, S., Mukhopadhyay, S., et al. (2015). K2P channel gating mechanisms revealed by structures of TREK-2 and a complex with Prozac. Science *347*, 1256–1259.

73. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

74. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

75. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

76. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

77. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

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

79. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. *33*, 1870–1874.

80. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Current Biology
Article

CellPress

81. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

82. Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012). Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front. Genet. 3, 35.

83. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

84. Kawabata, T. (2016). HOMCOS: an updated server to search and model complex 3D structures. J. Struct. Funct. Genomics 17, 83–99.

85. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics 54, 5.6.1–5.6.37.

86. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360.

87. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

88. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

89. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

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

91. Agarwala, R.; NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44 (D1), D7–D19.

92. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. 43, D222–D226.

93. Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 175–182.

94. Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol. Biol. 1079, 105–116.

95. Wintersinger, J.A., and Wasmuth, J.D. (2015). Kablammo: an interactive, web-based BLAST results visualizer. Bioinformatics 31, 1305–1306.

96. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

97. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191.

98. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.

99. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645.

100. Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics Chapter 2. Unit 2.3.

101. Gao, Y., Wang, S.Y., Luo, J., Murphy, R.W., Du, R., Wu, S.F., Zhu, C.L., Li, Y., Poyarkov, A.D., and Nguyen, S.N. (2012). Quaternary palaeoenvironmental oscillations drove the evolution of the Eurasian Carassius auratus complex (Cypriniformes, Cyprinidae). J. Biogeogr. 39, 2264–2278.

102. Omori, Y., Zhao, C., Saras, A., Mukhopadhyay, S., Kim, W., Furukawa, T., Sengupta, P., Veraksa, A., and Malicki, J. (2008). Elipsa is an early determinant of ciliogenesis that links the IFT particle to membrane-associated small GTPase Rab8. Nat. Cell Biol. 10, 437–444.

103. Petit, J., David, L., Dirks, R., and Wiegertjes, G.F. (2017). Genomic and transcriptomic approaches to study immunology in cyprinids: What is next? Dev. Comp. Immunol. 75, 48–62.

104. Izsvák, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., and Hackett, P.B. (1999). Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol. Evol. 48, 13–21.

105. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45 (D1), D183–D189.

106. Zhang, Y., Liu, J., Peng, L., Ren, L., Zhang, H., Zou, L., Liu, W., and Xiao, Y. (2017). Comparative transcriptome analysis of molecular mechanism underlying gray-to-red body color formation in red crucian carp (Carassius auratus, red var.). Fish Physiol. Biochem. 43, 1387–1398.

107. Hoekstra, H.E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97, 222–234.

108. Seberg, H.E., Van Otterloo, E., Loftus, S.K., Liu, H., Bonde, G., Sompallae, R., Gildea, D.E., Santana, J.F., Manak, J.R., Pavan, W.J., et al. (2017). TFAP2 paralogs regulate melanocyte differentiation in parallel with MITF. PLoS Genet. 13, e1006636.

109. van Eeden, F.J., Granato, M., Schach, U., Brand, M., Furutani-Seiki, M., Haffter, P., Hammerschmidt, M., Heisenberg, C.P., Jiang, Y.J., Kane, D.A., et al. (1996). Genetic analysis of fin formation in the zebrafish, Danio rerio. Development 123, 255–262.

110. Iovine, M.K., Higgins, E.P., Hindes, A., Coblitz, B., and Johnson, S.L. (2005). Mutations in connexin43 (GJA1) perturb bone growth in zebrafish fins. Dev. Biol. 278, 208–219.

111. Hama, H., Kurokawa, H., Kawano, H., Ando, R., Shimogori, T., Noda, H., Fukami, K., Sakaue-Sawano, A., and Miyawaki, A. (2011). Scale: a chemical approach for fluorescence imaging and reconstruction of transparent mouse brain. Nat. Neurosci. 14, 1481–1488.

112. Varshney, G.K., Carrington, B., Pei, W., Bishop, K., Chen, Z., Fan, C., Xu, L., Jones, M., LaFave, M.C., Ledin, J., et al. (2016). A high-throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish. Nat. Protoc. 11, 2357–2375.

113. Kawakami, K., Abe, G., Asada, T., Asakawa, K., Fukuda, R., Ito, A., Lal, P., Mouri, N., Muto, A., Suster, M.L., et al. (2010). zTrap: zebrafish gene trap and enhancer trap database. BMC Dev. Biol. 10, 105.

114. Wada, H., Ghysen, A., Asakawa, K., Abe, G., Ishitani, T., and Kawakami, K. (2013). Wnt/Dkk negative feedback regulates sensory organ size in zebrafish. Curr. Biol. 23, 1559–1565.

115. Asakawa, K., Suster, M.L., Mizusawa, K., Nagayoshi, S., Kotani, T., Urasaki, A., Kishimoto, Y., Hibi, M., and Kawakami, K. (2008). Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish. Proc. Natl. Acad. Sci. USA 105, 1255–1260.

116. Furutani, K., Ohno, Y., Inanobe, A., Hibino, H., and Kurachi, Y. (2009). Mutational and in silico analyses for antidepressant block of astroglial inward-rectifier Kir4.1 channel. Mol. Pharmacol. 75, 1287–1295.

117. Watanabe, M., Sawada, R., Aramaki, T., Skerrett, I.M., and Kondo, S. (2016). The Physiological Characterization of Connexin41.8 and Connexin39.4, Which Are Involved in the Striped Pattern Formation of Zebrafish. J. Biol. Chem. 291, 1053–1063.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Blood & Cell Culture DNA Maxi Kit	QIAGEN	Catalog number:13362
TRIzol reagent	Thermo Fisher Scientific	Catalog number:10296028
Agencourt AMPure XP	Beckman Coulter	Catalog number:A63881
TruSeq Stranded mRNA Library Prep Kit	Illumina	Catalog number:20020594
mMessage mMachine transcription kit	Thermo Fisher Scientific	Catalog number:AM1344
PrimeScript II Reverse Transcriptase	Takara	Catalog number:2690A
SYBR GreenER qPCR SuperMix	Thermo Fisher Scientific	Catalog number:11762100
Deposited Data		
Genome sequencing and RNA-seq reads	This paper	DRA: DRA008171
RNA-seq reads of seven tissues	[19]	SRA: SRP154139
Nucleotide sequence of the gFV-1	This paper	GenBank: LC471609
Goldfish genome sequence and annotation	Goldfish Genome Project	https://research.nhgri.nih.gov/goldfish/
DrFV-2	[69]	GenBank: CAAK05053864.1: 11,473–23,439
Tgf2	[70, 71]	GenBank: HM146132.1: 1453–2091
Structure of human K2P10.1	[72]	PDB: 4BW5
Genome sequencing and RNA-seq reads	This paper	DRA: DRA008171
Yellow River carp genome sequence and annotation	[27]	Genome Warehouse: GWHAATB00000000
Software and Algorithms		
Trimmomatic v0.38	[73]	http://www.usadellab.org/cms/?page=trimmomatic
GATK v4.0.4.0	[74]	https://software.broadinstitute.org/gatk/
SnpEff v4.2	[75]	http://snpeff.sourceforge.net/SnpEff.html
ADMIXTURE v1.3.0	[76]	http://software.genetics.ucla.edu/admixture/
R v3.4.3	The R Project for Statistical Computing	https://www.r-project.org
PLINK v1.90b4.5	[77]	https://www.cog-genomics.org/plink/
RAxML v8.2.12	[78]	https://github.com/stamatak/standard-RAxML
MEGA7	[79]	http://evolgen.biol.se.tmu.ac.jp/MEGA/
Neighbor-joining method	[80]	N/A
Pooled heterogeneity	[30, 31]	N/A
Gene ontology analysis	[81]	N/A
SnpSift v4.2	[82]	http://snpeff.sourceforge.net/SnpSift.html
IGV	[83]	https://software.broadinstitute.org/software/igv/
HOMCOS	[84]	http://homcos.pdbj.org/?LANG=ja
Modeler v9.17	[85]	https://salilab.org/modeller/
PyMOL v1.7.4.5.	N/A	https://pymol.org/2/
HISAT2 v2.1.0	[86]	http://ccb.jhu.edu/software/hisat2/index.shtml
HTSeq v0.8.0	[87]	https://htseq.readthedocs.io/en/release_0.8.0/
edgeR v3.16.5	[88]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Trinity v2.4.0	[89]	https://github.com/trinityrnaseq/trinityrnaseq
BLASTN v2.6.0	[90]	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
ORF Finder	[91]	https://www.ncbi.nlm.nih.gov/orffinder/

Current Biology
Article

CellPress

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Conserved Domains database	[92]	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
TMHMM Server v.2.0	[93]	http://www.cbs.dtu.dk/services/TMHMM/
Clustal Omega v1.2.4	[94]	http://www.clustal.org/omega/
Kablammo	[95]	http://kablammo.wasmuthlab.org/
trimAl v1.4	[96]	http://trimal.cgenomics.org/
Jalview	[97]	https://www.jalview.org/
EMBOSS	[98]	http://emboss.sourceforge.net/download/
CIRCOS v0.69-9	[99]	http://circos.ca/
CLUSTALW v2.1	[100]	https://www.genome.jp/tools-bin/clustalw

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to, and will be fulfilled by the Lead Contact. Yoshihiro Omori (omori4@nagahama-i-bio.ac.jp). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

All procedures were approved by the Institutional Safety Committee on Recombinant DNA Experiments (approval ID 04220) and the Animal Experimental Committees of the Institute for Protein Research (approval ID 29-01-0) at Osaka University and were performed in compliance with the institutional guidelines. All goldfish in this study were bred in Japan (Data S1C). All adult goldfish strains, except Albino Celestial and Albino Ranchu, were purchased from a provider, Meito Suien (https://remix-net.co.jp/). Details of the strains are presented in Data S1B and Figure S3Z'. Albino Celestials and Albino Ranchu were obtained from the Aichi Fisheries Research Institute, and wild goldfish were obtained from the Lake Kasumigaura and Lake Hinuma basins. We selected wild goldfish (C6 clade) following mitochondrial DNA sequence analysis [101]. Fertilized goldfish eggs were incubated at 26°C in fresh water. Three to 5 days post fertilization (dpf), the hatched goldfish larvae were fed brine shrimp (*Artemia*) twice daily. At 14 dpf, the goldfish were fed pellets daily. The water in the larvae and adult goldfish tanks was changed on a weekly and monthly basis, respectively. The maintenance and breeding of the zebrafish strains (Tuebingen) as well as embryonic development staging were performed as described previously [102].

METHOD DETAILS

Partitioning the goldfish genome into two homeologous subgenomes

Allopolyploidization is the doubling of chromosome sets following an interspecific hybridization of two progenitor species [103]. When the two progenitors exist as distinct species, the TEs in their genomes can be independently active. Therefore, the relicts of TEs that are specific to each progenitor indicate the descendent subgenomes in an allotetraploid genome [23]. From the Repeat-Masker output (https://research.nhgri.nih.gov/goldfish/), we calculated the subgenome-biased index (SBI), of which the formula is presented in Figure S1A. We manually verified the TE distributions (SBI > 0.6) and identified the TEs with subgenome-biased distributions.

Phylogenetic analysis of the subgenome-specific TEs

Sequences of *Tc1-1* (L-subgenome-specific TE) or *Mariner-12* (S-subgenome-specific TE) were obtained from the goldfish reference genome sequence using the RepeatMasker output. We used Clustal Omega to align the DNA sequences and a sequence of their corresponding zebrafish orthologs for both TEs. Poor alignments were trimmed by trimAl v1.4 [96]. The alignments were manually corrected using Jalview [97]. Then, we constructed neighbor-joining trees utilizing the Jukes–Cantor substitution model in MEGA7. These trees included multiple 'star' topologies, representing a single ancestral TE that was active at a defined time [23]. Only star topologies with > 5 copies were used for further analysis. For each 'star' topology, we computed a consensus sequence and calculated the Jukes–Cantor distances between the TEs and their respective consensus sequences using MEGA7.

As an example of the recently or currently active TE, we identified a subset of *angel* [104] in the goldfish reference genome by BLASTN with the internal inverted repeat region of Tgf2 (HM146132.1: 1453–2091) as a query sequence [70, 71]. Common carp *angel* was identified by the same approach. The phylogenetic analysis of *angel* was performed using the same approach as for *Tc1-1* and *Mariner-12*, except that the sequence of common carp *angel* was used as the outgroup.

Molecular evolutionary analysis of goldfish subgenomes

Lists of 1:2 orthologs between zebrafish and goldfish and the singleton genes in goldfish were obtained from another study [19]. We focused on 5,404 gene pairs located in goldfish LG1–LG50 and zebrafish chromosomes 1–25. For each goldfish ohnolog gene pair, the longest representative gene models were selected, and the amino acid sequences of the goldfish ohnologs were aligned to the zebrafish ortholog using Clustal Omega. The amino acid sequences in the multiple alignments of three sequences were then replaced with their corresponding coding sequences using tranalign from the EMBOSS software suite [98]. Each aligned goldfish ohnolog was compared independently with its zebrafish ortholog, and we calculated the ratio of the nonsynonymous substitution rate per nonsynonymous site (dN) to the synonymous substitution rate per synonymous site (dS) using MEGA7.

The RNA-seq data of seven tissues used for goldfish genome annotation (SRP154139) were aligned to the reference genome. The gene abundances were estimated as described above. The average reads per kilobase of exon per million mapped sequence reads (RPKM) was calculated for each tissue. To calculate the average RPKM of the eye, the RNA-seq data from the eyes of the three common goldfish in this study were included.

Comparative genome analysis between goldfish and common carp

The genome sequence and annotation of Yellow River carp (accession number: GWHAATB00000000) were downloaded from Genome Warehouse. To resolve the orthology between 50 chromosomes of goldfish and those of common carp, we analyzed ohnologs with gene coordinate information that has a relatively high reliability. In goldfish, we chose 5,404 pairs of ohnologs based on the following three criteria: (1) the number of ohnologs in the genome assembly was exactly two; (2) both ohnologs were located on the placed scaffolds; and (3) the two ohnologs were located on homeologous chromosomes. For common carp, we downloaded the 8,291 ohnolog pairs from a previous report [27]. The protein-coding sequences of these ohnologs were compared between goldfish and common carp by BLASTN, and reciprocal best hits were identified. The results were visualized with CIRCOS v0.69-9 [99]. The nucleotide sequences of hAT-N91 and Mariner-N17 were searched against the genome sequence of Yellow River carp by BLASTN. Based on the genome annotation of common carp, 44,626 protein-coding sequences were retrieved from the common carp genome. The coding sequences of the 2,099 singleton genes of goldfish [19] were subjected to BLASTN against the 44,626 protein-coding sequences of common carp. The 1,997 singleton genes of common carp [27] were subjected to BLASTN against the 44,650 protein-coding sequences of goldfish, which are located in ohnolog blocks, as reported previously by us [19].

Purification of genomic DNA and total RNA from the goldfish tissues

The goldfish were anesthetized using tricaine, the caudal fins were dissected, and high molecular weight genomic DNA was purified using TissueLyser II (QIAGEN) and Blood & Cell Culture DNA Maxi Kit (QIAGEN). Following agarose gel electrophoresis and ethidium bromide staining, 20-kb bands of purified genomic DNA were identified. For RNA-seq analysis, total RNA, obtained from the dissected eyeballs of goldfish that were anesthetized with tricaine, was purified using TRIzol reagent (Thermo Fisher Scientific).

Whole-genome sequencing and RNA-seq of the goldfish strains and wild goldfish

For the whole-genome sequencing of goldfish strains, genomic DNA was sheared to an average size of 600 bp using an M220 Focused-ultrasonicator (Covaris). Paired-end libraries were prepared with a TruSeq DNA PCR-Free Library Prep Kit (Illumina) and then purified using Agencourt AMPure XP (Beckman Coulter). The DNA concentration of the libraries was measured on an Agilent 2100 Bioanalyzer (Agilent Technologies), and the libraries were sequenced on HiSeq 2500 sequencers (Illumina) with a read length of 250 bp.

For the RNA sequencing, 4 μg of total RNA from each goldfish strain was used for the library preparation. The sequencing libraries were constructed using the TruSeq Stranded mRNA Library Prep Kit (Illumina). The cDNA sample was used following 15 cycles of PCR amplification without size selection. The libraries were sequenced on Illumina HiSeq 2500 sequencers. The read length was 100 bp for all libraries.

Read trimming and mapping

The raw reads obtained from the whole-genome sequencing data were trimmed using Trimmomatic v0.38 (LEADING:30 TRAILING:30 SLIDINGWINDOW:4:25 MINLEN:50) [73]. The goldfish reference genome (carAur01) that was soft-masked by RepeatMasker (http://www.repeatmasker.org), obtained from the Goldfish Genome Project website (https://research.nhgri.nih.gov/goldfish/), was used as a mapping reference. The filtered reads were aligned to the goldfish reference genome by our in-house program, which was set at > 92% sequence identity with the reference sequence. Multi-mapping, where it is difficult to map a sequencing read to a particular position, is often the cause of incorrect genotyping. A read was considered as a multi-hit read if the difference between the first and second highest sequence identities between the read and the reference genome was 2% or less. Reads other than multi-hit reads were regarded as uniquely mapped reads. Uniquely mapped pair reads were analyzed further.

Variant calling, filtering, and annotations

SNVs and indels of goldfish individuals were called using GATK v4.0.4.0 software [74]. All calls from individuals were combined into one dataset. Following the removal of variants with a read depth < 3, there were 63,347,761 variants (48,062,591 SNVs and 15,285,170 small indels) remaining. Variants with a call rate < 100% and minor allele frequency < 1% were filtered out. To perform variant annotations, the goldfish genome annotation in General Feature Format was downloaded from the Goldfish Genome Project

Current Biology
Article

** CellPress**

website, and the variants were annotated with SnpEff v4.2 [75]. We retained 9,271,048 variants (7,564,197 SNVs and 1,706,851 indels) as the initial dataset for further analysis.

Estimation of the subgenome-specific SNV and indel rates between the goldfish strain genomes and the reference genome

To strictly estimate the subgenome-specific SNV and indel rates between the goldfish strain genomes and the reference genome, the individual variant calls, in genomic variant call file format, were subjected to filtering. Regions with a read depth of ≥ 2 reads on both strands and within 2.5 SD of the mean depth of each sample were retained. Regions with $ZH_p < -1.5$ were excluded. Continuous regions > 100 bp (SNVs) or > 200 bp (indels) were retained. To estimate the SNV rates, 20-bp regions around the indels were excluded to filter false-positives. After these filtering steps, we calculated the SNV and indel rates for both subgenomes.

Admixture analysis and grouping of the goldfish strains

The initial dataset of 7,564,197 SNVs was subjected to filtering. Only SNVs mapped in LG1–LG50 were retained for the admixture analysis. Following exclusion of the SNVs mapped in unplaced scaffolds, 6,458,046 SNVs remained. Admixture analysis was conducted using ADMIXTURE v1.3.0 software [76] with cross-validation for K values from 1 through 15. We classified 48 goldfish individuals into three groups according to the proportion of the individual's genome from the inferred ancestral populations at K value = 3. We designated these groups China, Ranchu, and Edo. The admixture analysis results were visualized using R v3.4.3 (https://www.r-project.org).

Principal component analysis (PCA)

We performed PCA using PLINK v1.90b4.5 software, based on the variance-standardized relationship matrix [77], to analyze the goldfish population structure. The distance matrix was computed based on the 6,458,046 SNVs mapped in LG1–LG50. The PCA results were plotted using R.

Phylogenetic analysis of the goldfish strains and the wild goldfish

To conduct a phylogenetic analysis of the goldfish individuals, we calculated pairwise genome-wide identical-by-state (IBS) distances based on the 6,458,046 SNVs mapped in LG1–LG50 by PLINK. The maximum-likelihood tree was constructed with 500 bootstrap replicates for node support using RAxML v8.2.12 [78]. Based on the pairwise distance matrix (1–IBS), a neighbor-joining tree [80] was constructed with 500 bootstrap replicates for node support using MEGA7 [79].

Fst analysis

To calculate Fst, we used the 6,458,046 SNVs mapped in LG1–LG50. The China, Ranchu, and Edo groups were included in this analysis. For each of the above groups, Fst values were calculated between individuals belonging to the group and the rest in 40-kb windows sliding 10 kb at a time. These calculations were performed using PLINK. If the Fst value of one group in one window was greater than 0.05 and the Fst values of the other two groups in the same window were less than 0.05, the first of the three groups was regarded as significantly differentiated from the other two groups.

Pooled heterogeneity of the goldfish strains

To identify the regions with a high degree of fixation in the goldfish strains, pooled heterogeneity (H_p) was calculated in 40-kb windows sliding 10 kb at a time [31, 32]] using the dataset of 46 goldfish individuals representative of all 27 strains. We calculated the Z-transformed distribution of H_p (ZH_p) and extracted the windows at the extreme ends of the distribution by applying a cutoff of $ZH_p < -1.5$, leaving only the regions with continuous windows > 300 kb remaining. We identified 2,020 genes in these regions and performed gene ontology (GO) analysis using GO Enrichment Analysis [105].

Identification of strain-specific variants

We analyzed homozygous strain-specific variants (SSVs) for those strains where genome sequencing analysis was performed for > 2 individuals (18 strains). First, SSVs were identified using SnpSift [82]. If > 2 SSVs were identified within a 50-kb region, they were recognized as a cluster. We defined an SSV-enriched region (SSVR) as a cluster > 100 kb that contained > 5 SSVs.

Phylogenetic analysis of the goldfish strains, wild goldfish, crucian carp, and common carp

Using a publicly available dataset of restriction site-associated DNA sequencing (RAD-seq) of closely-related goldfish species (NCBI Sequence Read Archive, accession number: SRP063043), the RAD-seq reads were aligned to the goldfish reference genome, and 86,404 SNVs were identified as described above. This dataset was combined with our genome sequencing dataset. We selected the genomic positions with a 95% genotyping call rate among all individuals and used these to construct a neighbor-joining phylogenetic tree as described above.

Genome-wide association study (GWAS)

The dataset was divided into cases or controls based on the phenotypes (Data S1B and S1I), namely, telescope-eye, long-tail, twin-tail, dorsal fin loss, hood, shortened body, and calico. Single-marker chi-square association tests were performed by PLINK using the

initial variants dataset (9,271,048 variants). *P*-values were adjusted by Bonferroni correction, and those $< 1 \times 10^{-10}$ were deemed significant. The significantly associated regions were manually verified from the aligned sequencing reads against the goldfish reference genome using Integrative Genomics Viewer (IGV) [83].

Homology modeling of the goldfish Kcnk5b structure
We used the HOMCOS server [84] to search for a template for homology modeling of the goldfish Kcnk5b structure and identified human K2P10.1 (PDB ID: 4bw5) for use as a template. A model of the goldfish Kcnk5b structure was built with Modeler [85] and visualized with PyMOL v1.7.4.5. (https://pymol.org/2/).

RNA-seq analysis
RNA-seq reads were trimmed as described above and then aligned to the goldish reference genome using HISAT2 v2.1.0 [86]. Reads aligned concordantly exactly one time were used to avoid redundancy. The gene abundance was estimated using HTSeq v0.8.0 [87].

Identification of gFV-1 transposable element insertion in lrp2aL of Black Telescope-eye goldfish
De novo transcriptome assembly of the RNA-seq data for *lrp2aL* was obtained using Trinity v2.4.0 [89] with the default parameters. Then, using BLASTN v2.6.0 [90] with the default parameters, we searched for the sequence containing *lrp2aL* exon 45 from the contig sequences produced by Trinity. We determined the sequence of a 13-kb PCR fragment amplified from *lrp2aL* intron 45 in the genome of Black Telescope-eye goldfish and found that this insertion encoded a foamy-like endogenous retrovirus. We named this transposable element (TE) gFV-1. Using BLASTN with the gFV-1 sequence as a query, we identified 15 sequences that shared at least a 5-kb sequence portion of gFV-1 with > 95% sequence identity to the goldfish reference genome. An ORF analysis using ORF Finder [91], classified these sequences into two groups based on whether they contained reverse transcriptase (RT). The consensus sequences of putative autonomous and non-autonomous gFV-1 were named gFV-1_PA and gFV-1_NA, respectively. These sequences were annotated with NCBI's Conserved Domains database [92]. Transmembrane regions were predicted with TMHMM Server v.2.0 [93].

Phylogenetic analysis of gFV-1_NA, gFV, and their derivatives
The *pol* sequences from gFV-1, gFV-1_NA, and their derivatives were identified in the reference goldfish genome. The *pol* protein sequences of gFV-1 and zebrafish DrFV-2 (CAAK05053864.1: 11,473–23,439) [69] were aligned in Clustal Omega [94]. A maximum-likelihood phylogenetic tree was constructed, and its reliability was assessed with 500 bootstrap replicates.

Synteny analysis of the lrp2aL locus in LG9 and LG34
The alignment of the genomic sequence containing *lrp2aL* (7,845,000–8,162,000) on LG9 against the corresponding sequence (3,015,000–3,211,000) on LG34 was generated using BLASTN and visualized with Kablammo [95].

Candidate gene list for pigmentation and fin shape
Based on previous literature, 143 candidate genes for body-color-related genes were selected (Data S1N) [106–108], whereas five candidate genes (*kcnk5b*, *kcnh2a*, *rpz*, *slc12a7*, and *cx43*) for fin shape formation were selected based on previous literature [46, 57, 109, 110].

Skeletal staining of goldfish
Skeletal staining using alizarin red and alcian blue was performed as previously described with some modifications [32]. Briefly, goldfish were anesthetized with tricaine, fixed with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) at 4°C overnight, washed with 70% ethanol, and stained with alcian blue staining solution (0.1% alcian blue in 75% ethanol/25% acetic acid) for 2–4 h. Then, the specimens were washed with 70% ethanol and stained with alizarin red solution (0.1% alizarin red in 95% ethanol) for 2–4 h, washed with 70% ethanol, and cleared with ScaleA2 [111].

CRISPR/Cas9-mediated target mutagenesis in zebrafish
The gRNA synthesis and injection into zebrafish embryos were performed as previously described with some modifications [112]. Briefly, gRNA was synthesized using a MEGAshortscript T7 Transcription Kit (Thermo Fisher Scientific) with PCR products containing the target-specific sequence. The primers used for the amplification are detailed in Table S2. Cas9 3NSL enzyme (Integrated Data Technologies) and gRNA were injected into zebrafish embryos at the single-cell stage using a FemtoJet microinjector (Eppendorf), Femtotip II microcapillaries (Eppendorf), and 1.5% agarose molds. We injected CRISPR mix reagents into 200 embryos of each group. From the surviving larvae with a straight body, we randomly selected 13 (control), 11 (exon 2), and 12 (exon 33) larvae from each group. We took photographs of these individuals using stereomicroscopy and measured the body length. For genotyping, we amplified DNA fragments by PCR using the primers detailed in Table S2. For the genotyping of mixed sequences, we subcloned PCR products into the pCR-BluntII-TOPO (Thermo Fisher Scientific) vector, and sequenced.

Partial inhibition of Wnt signaling by ectopic expression of Dkk1
We used a Gal4 enhancer trap line, *hspGFF55B* [113], and transgenic lines *UAS:dkk1a-rfp* [114] and *UAS:egfp* [115]. The zebrafish larvae were viewed under a STeREO Lumar. V12 microscope (Carl Zeiss).

Current Biology
Article

ɔ **CellPress**

Voltage clamp recording

Goldfish wild-type *kcnk5b* (*gfkcnk5b*) and g*fkcnk5b* (V165E) were each cloned in pGEM-HeFx plasmids [116]. Following linearization with a restriction enzyme, each cRNA was prepared using a mMessage mMachine transcription kit (Ambion). *Xenopus* oocytes were collected as described previously [117] and injected with 0.2 ng of each cRNA. H_2O was injected as the negative control. Following overnight incubation at 18°C, the transmembrane current was measured in ND96 (+) solution (93.5 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2$, 2 mM $MgCl_2$, and 5 mM HEPES [pH 7.5, adjusted with NaOH]) using an iTEV90 multi electrode clamp amplifier (HEKA). The cells were initially clamped at −80 mV and then subjected to 500 ms voltage steps from −100 mV to +60 mV in 20 mV increments.

Quantitative RT-PCR (qRT-PCR) analysis

We used TRIzol reagent (Thermo Fisher Scientific) to purify total RNA from the goldfish eyeballs or embryos anesthetized with tricaine. The cDNA was synthesized using PrimeScript II Reverse Transcriptase (Takara). qRT-PCR was performed using SYBR GreenER qPCR SuperMix (Thermo Fisher Scientific) on a Thermal Cycler Dice Real Time System Single MRQ TP870 (Takara). Quantification was carried out using Thermal Cycler Dice Real Time System v2.0 software (Takara). The primer sequences are detailed in Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of single comparisons were performed using paired or unpaired Student's t test,. Data are reported as mean ± SD using GraphPad Prism version 6.04 (GraphPad Software). The analyzed number of samples is indicated in the figure legends. Asterisks indicate significance values as follows: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

DATA AND CODE AVAILABILITY

The accession numbers for the raw reads reported in this paper are the DDBJ Sequence Read Archive (DRA): DRR172174–DRR172230. The accession number for the nucleotide sequence of the gFV-1 is DDBJ: LC471609.