Check for updates

OPEN

Large-scale sequencing of flatfish genomes provides insights into the polyphyletic origin of their specialized body plan

Zhenming Lü [1,7], Li Gong[1,7], Yandong Ren[2,7], Yongjiu Chen[1,7], Zhongkai Wang [2,7], Liqin Liu[1,7], Haorong Li[2,7], Xianqing Chen[2], Zhenzhu Li[2], Hairong Luo[3], Hui Jiang[1], Yan Zeng[4], Yifan Wang[1], Kun Wang[2], Chen Zhang[2], Haifeng Jiang[5], Wenting Wan[2], Yanli Qin[2], Jianshe Zhang[1], Liang Zhu[4], Wei Shi[3], Shunping He[5], Bingyu Mao[4], Wen Wang [2,4,6,8 ✉], Xiaoyu Kong [3,8 ✉] and Yongxin Li [2,4,8 ✉]

The evolutionary and genetic origins of the specialized body plan of flatfish are largely unclear. We analyzed the genomes of 11 flatfish species representing 9 of the 14 Pleuronectiforme families and conclude that Pleuronectoidei and Psettodoidei do not form a monophyletic group, suggesting independent origins from different percoid ancestors. Genomic and transcriptomic data indicate that genes related to WNT and retinoic acid pathways, hampered musculature and reduced lipids might have functioned in the evolution of the specialized body plan of Pleuronectoidei. Evolution of Psettodoidei involved similar but not identical genes. Our work provides valuable resources and insights for understanding the genetic origins of the unusual body plan of flatfishes.

The colonization of the seafloor is one of the most important events in evolutionary history, which led to an explosive radiation and large-scale morphological diversification of marine phyla[1,2]. Flatfishes are one of the most successful groups of seafloor colonizers and have evolved a specialized morphology that is unique in teleosts. Such morphological innovations include a flat and thin body plan that facilitates embedding into substrates[3], an asymmetrical body axis, mostly represented by one eye migrating to the contralateral side of the skull for gaining binocular vision, which ensures improved success of preying[4], and modified median and paired fins that coordinate together to enable flexible over-substrate 'fin-feet' walking[5,6]. The body plan exhibited by flatfishes reflects morphological trade-offs to facilitate embedding, predation and maneuvering behaviors adapted to their over-substrate dwelling lifestyle[4,6]. However, the genetic basis of such morphological adaptations in flatfishes has remained largely unknown since the time of Darwin[7,8].

Some progress has been made concerning the evolutionary origin and the morphological adaptations of flatfishes in recent years. Current views support the origin of flatfishes among basally diverging percoids[9–11]. Despite this progress, there is still disagreement regarding when and how flatfishes diverged from their ancestors. An unanswered question is whether the flatfishes (particularly Pleuronectoidei and Psettodoidei, the only two suborders of Pleuronectiformes) share a monophyletic origin[9,12–14]. This has been difficult to address due to limitations in providing a solid evolutionary framework for understanding the genetic basis of the morphological adaptations of flatfishes; largely because the

supposed polyphyletic origin may predict differed genetic mechanisms for their morphological adaptations. The early exploration of the genetic origin of the specialized morphology of flatfishes was started by Inui et al.[15], and continued by Hashimoto et al.[16,17] and Suzuki et al.[18], but the results varied from either a NODAL or thyroid hormone (TH) regulation of their asymmetrical body plan. Shao et al.[19] were the first to elaborate on this topic by applying a genomic framework and providing evidence for retinoic acid (RA) and TH involvement in the regulation of body plan asymmetry of flatfishes. However, all these studies mainly focused on the asymmetric body plan in flatfishes under the framework of only one or two flatfish species, while the genetic basis of a wider spectrum of morphological adaptations (for example, body-plan flatness, body and eye asymmetry and fin modification) in the whole flatfish group remains to be explored from a systematic evolutionary perspective.

In the present study, we assembled genomes of eight species de novo (*Trinectes maculatus*, *Chascanopsetta lugubris*, *Brachirus orientalis*, *Paraplagusia blochii*, *Colistium nudipinnis*, *Pseudorhombus dupliocellatus*, *Platichthys stellatus* and *Psettodes erumei*) representing most of the major extant clades (8 of 14 families, including the sole family in Psettodoidei and 7 families in Pleuronectoidei) of Pleuronectiformes and two closely related species of Perciformes with regular body plans (*Toxotes chatareus* and *Polydactylus sextarius*). Combined with three previously published genomes of flatfish species, which added one more family, and 80 transcriptomes (including 72 from three tissues of *Paralichthys olivaceus*; 4 from two tissues of *Platichthys stellatus*; 2 from two tissues of *Toxotes*

¹National Engineering Laboratory of Marine Germplasm Resources Exploration and Utilization, Zhejiang Ocean University, Zhoushan, China. ²School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, China. ³CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China. ⁴State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. ⁵Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China. ⁶Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. ⁷These authors contributed equally: Zhenming Lü, Li Gong, Yandong Ren, Yongjiu Chen, Zhongkai Wang, Liqin Liu, Haorong Li. ⁸These authors jointly supervised this work: Wen Wang, Xiaoyu Kong, Yongxin Li. ✉e-mail: wwang@mail.kiz.ac.cn; xykong@scsio.ac.cn; yxli28science@sina.com

chatareus and 2 from two tissues of *Polydactylus sextarius*) that we generated, we systematically studied: (1) the phylogeny of flatfishes, which provides an evolutionary framework for better understanding the genetic adaptation of flatfishes; and (2) genes that experienced significant alterations, to gain insights into the genetic basis underlying the unusual body plan of flatfishes.

Results

Genome assembly and annotation. Using whole-genome sequencing strategies, we generated more than two terabytes (Tb) of sequencing data (Supplementary Tables 1–13 and Supplementary Notes 1–5) and de novo-assembled genomes of the ten species indicated above (Fig. 1a, Supplementary Tables 14–23 and Supplementary Notes 6 and 7). Among them, three species with controversial phylogenetic status, including *Psettodes erumei* (Pleuronectiformes), *Toxotes chatareus* (Perciformes) and *Polydactylus sextarius* (Perciformes), were sequenced using a Nanopore platform (Supplementary Tables 1 and 12 and Supplementary Notes 1, 3 and 4). Hi-C data analysis supported the generation of chromosome-level genome assemblies for three species: *Platichthys stellatus*, *Toxotes chatareus* and *Polydactylus sextarius* (Fig. 1b–d, Supplementary Tables 24–29, Supplementary Figs. 1–3 and Supplementary Note 8). All ten assembled genomes possess high continuity and accuracy as indicated by the N50 length (64.40 kilobases (kb)–25.10 megabases (Mb); Supplementary Tables 14–26 and Supplementary Notes 6–8), genome integrity (Supplementary Table 30, Supplementary Fig. 4 and Supplementary Note 9), BUSCO scores (93.6–99.1%; Supplementary Tables 31–40 and Supplementary Note 10), read mapping ratios (94.64–99.87%; Supplementary Table 41 and Supplementary Note 10) and transcript mapping ratios (95.06–99.32%; Supplementary Tables 42–47 and Supplementary Note 10). The quality of the chromosome-level assemblies was also demonstrated by the good genome synteny (Fig. 1e, Supplementary Figs. 5–9 and Supplementary Note 11). The assembled genome sizes range from 399.64 Mb (*Pseudorhombus dupliocellatus*) to 643.91 Mb (*Paralichthys olivaceus*) (Fig. 1f, Supplementary Tables 14–30 and Supplementary Notes 6–9). After masking repetitive sequences (Supplementary Tables 48–68, Supplementary Fig. 10 and Supplementary Notes 12 and 13), these genomes were predicted to contain ~20,000 protein-coding genes (Supplementary Tables 69–79 and Supplementary Notes 14 and 15), which share similar gene structures to the published genomes (Extended Data Fig. 1 and Supplementary Note 14).

Polyphyletic origin of flatfishes. By combining our ten de novo-assembled genomes with eight published genome sequences from teleost species of *Cynoglossus semilaevis*, *Paralichthys olivaceus*, *Scophthalmus maximus*, *Larimichthys crocea*, *Labrus bergylta*, *Oreochromis niloticus*, *Oryzias latipes* and *Danio rerio* (see Supplementary Note 16), we reconstructed the phylogeny of flatfishes using concatenated sequences of coding sequence (CDS) (codon1 + 2 + 3, GTRGAMMA model; codon1 + 2, GTRGAMMA model) and 4dTV (fourfold degenerate synonymous site, GTRGAMMA model) derived from 1,693 single-copy genes (Supplementary Figs. 11–15 and Supplementary Notes 16 and 17). We further constructed the species tree under the coalescent model[20,21]. Our results consistently show that *Psettodes erumei* of

suborder Psettodoidei forms one clade with the two Perciformes species with regular body plan, *Toxotes chatareus* and *Polydactylus sextarius*, and species of suborder Pleuronectoidei form its sister clade (Fig. 2a, Supplementary Figs. 16 and 17 and Supplementary Note 17). The observation that, in both gene trees and species trees, *Psettodes erumei* is clustered with nonflatfish Perciformes rather than with Pleuronectoidei species provides strong support for the independent origins of Pleuronectoidei and Psettodoidei. Alternatively, it is also possible that they had a monophyletic origin but secondarily lost their traits independently, in *Toxotes chatareus* and *Polydactylus sextarius*. However, considering that *Psettodes erumei* has also been observed to show affinity to *Sphyraena argentea*, *Centropomus armatus*, *Coryphaena hippurus*, *Nematistius pectoralis* and many other perciforme species rather than Pleuronectoidei in multiple previous phylogenetic studies[22,23], this scenario is less likely, because multiple independent losses in many species along a lineage are less likely according to the parsimony principle of evolution[24]. We also analyzed mutations in body-plan-related genes, and conclude that many reverse mutations would need to have arisen if we assume secondary losses in *Toxotes chatareus* and *Polydactylus sextarius* (Supplementary Table 80 and Supplementary Note 17), which is less likely in molecular evolution[25]. Furthermore, reconstruction of ancestral chromosomes for the Pleuronectoidei and Psettodoidei lineages also shows that *Psettodes erumei* shares specific chromosome rearrangements with *Toxotes chatareus* and *Polydactylus sextarius*, rather than with Pleuronectoidei species, further supporting a polyphyletic origin for these two lineages (see Supplementary Note 17). Indeed, the morphological resemblance between Psettodoidei and percoids has long been noticed by several ichthyologists[26–28], and Psettodoidei were even once regarded as 'simply an asymmetric percoid'[9,26]. The morphological differentiation of Psettodoidei from Pleuronectoidei includes: (1) lack of skin folds around the eyes[29]; (2) posterior insertion of the dorsal fin[30]; (3) less extensive cranial asymmetry[14]; (4) presence of spinous rays in fins[31]; and (5) larger mouths with specialized teeth[31]. These phenotypical observations, combined with our results, provide strong support for a polyphyletic origin of flatfishes, with Psettodoidei and Pleuronectoidei, respectively, arising from two independent evolutionary events. To capture real evolutionary signals, we therefore split the previously known Pleuronectiformes into 'real flatfish Pleuronectoidei' (RFP) and 'flatfish-like Psettodoidei' (FLP) lineages in the following analyses.

Fast evolution in flatfishes. With fossil calibration, we estimated the emergence of RFP and FLP to be approximately 76.1 and 80.0 million years ago (Ma), respectively, in the late Cretaceous (Fig. 2a). Our time estimates are consistent with the calibration in a previous study using multiple nuclear loci[9], which is earlier than other estimations using mitochondrial or few nuclear loci[12,32]. The late Mesozoic to early Cenozoic period, which includes the Cretaceous, is known as the 'second age of fishes', marking the onset of major diversifications and morphological diversification of teleosts[33,34]. Guinot and Cavin[35] attributed such radiations to the combined effects of exceptionally high seawater temperatures, increasing sea levels and widespread epicontinental seas during the period[36,37]. The period of 80–75 Ma during the late Cretaceous experienced a peak of

Fig. 1 | Genome assembly and gene annotation of ten species. a, Geographical distribution of the species sequenced in this study. The dark green areas on the maps represent the global distribution regions of each species, which were acquired from the open source FishBase database (www.fishbase.se). The L, R and LR represent sequenced species that have left-side eyes, right-side eyes or one eye on each side of the body, respectively. **b–d**, Circos plot of distribution of the genomic elements in the species *Polydactylus sextarius* (**b**), *Toxotes chatareus* (**c**) and *Platichthys stellatus* (**d**). From outer to inner ring are the distributions of protein-coding genes, tandem repeats (TRP), long terminal repeats (LTR), short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE) and GC content, respectively. **e**, Genomic colinearity of three chromosome-level genomes. The number in the circles represents the chromosome identity for each species. **f**, Genome size statistics of these species. Diagram indicates the size of each type of element, including the coding regions, DNA elements, LINE, SINE, LTR and other genomic regions, in each species.

such global change[37]. We hypothesize that such fast seafloor spreading and the resulting explosion of epicontinental habitat may have facilitated the seafloor colonization and eventual origin of flatfishes.

Such a scenario also predicts selection pressure and faster evolutionary rate in RFP and FLP, since they experienced radical habitat transition from water column to seafloor. To test this hypothesis,
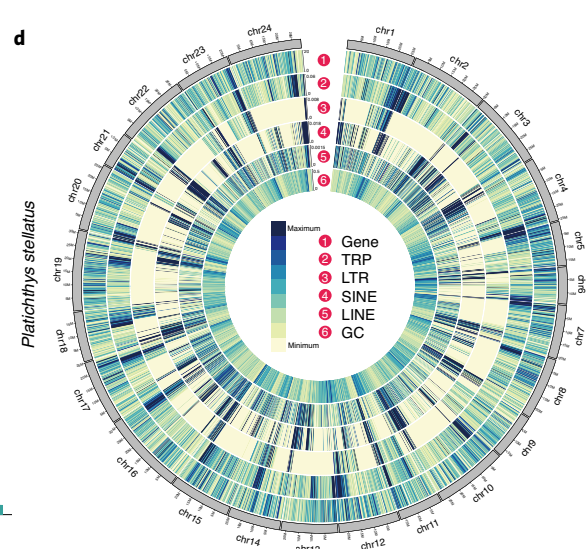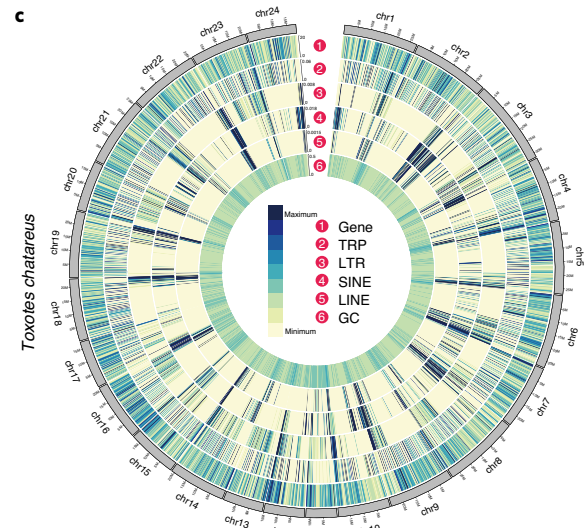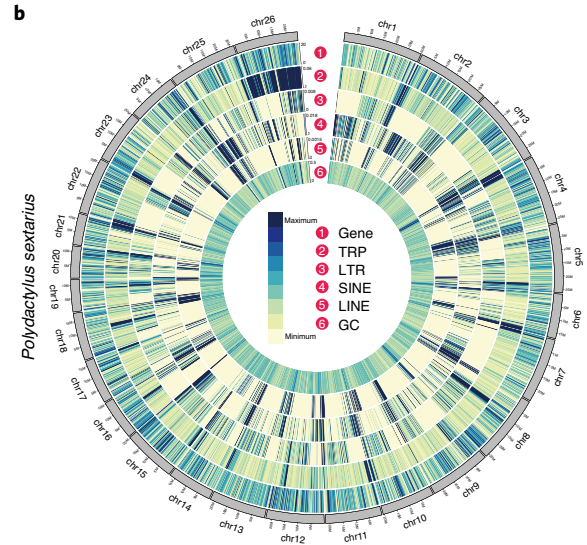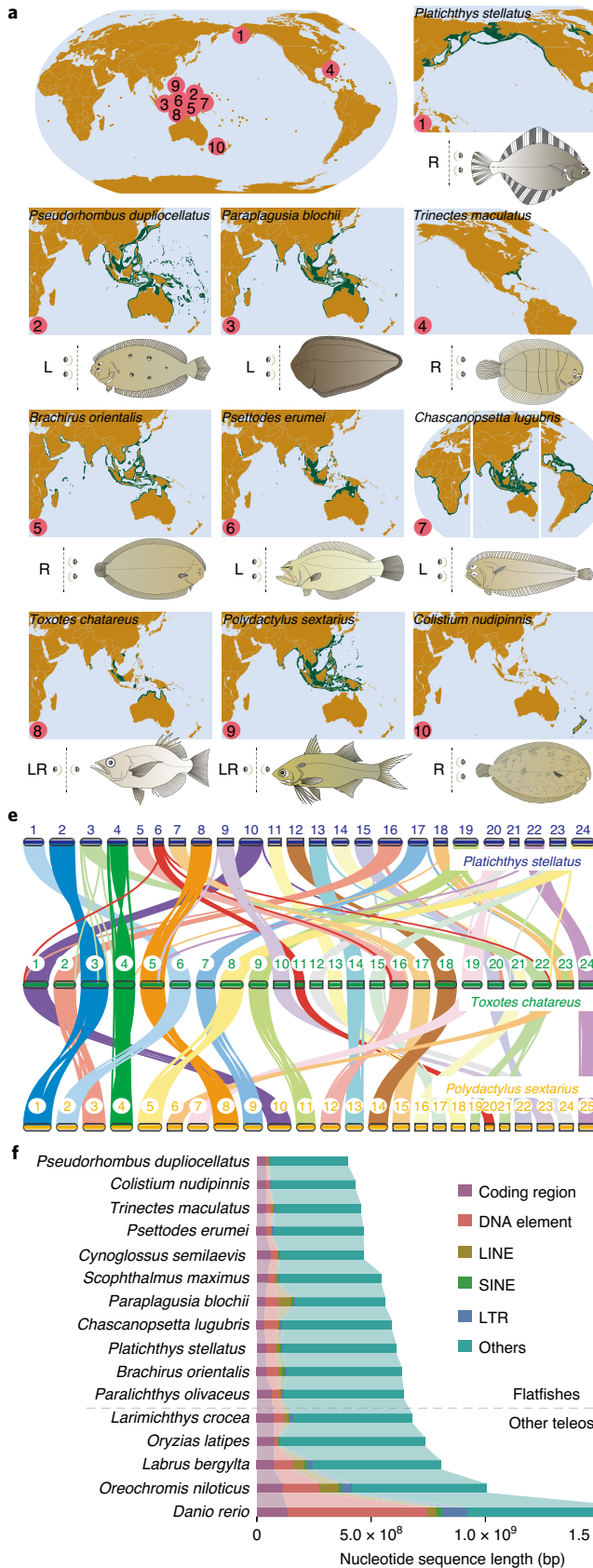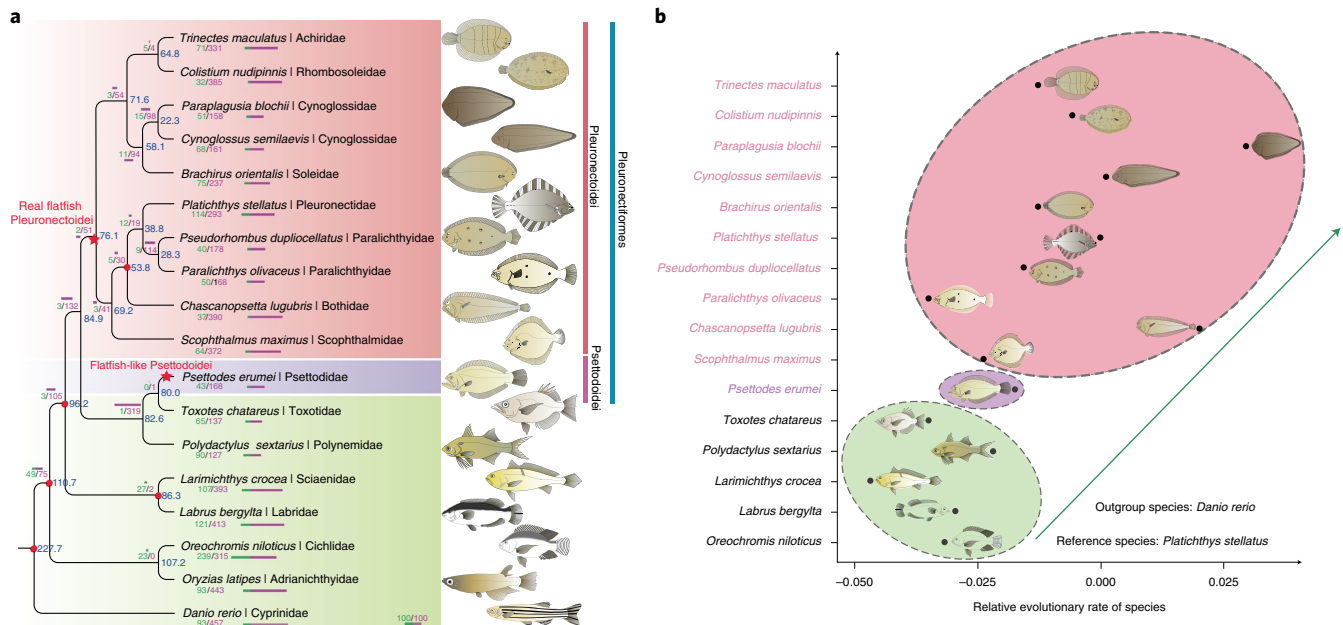
Fig. 2 | Polyphyletic origin and fast genome evolution of flatfishes. a, Phylogenetic relationship and divergence time among the flatfish species. The number in each node represents the divergence time among species and the red circle indicates the fossil record used for calibration in the node. The numbers with slashes, as well as the bars, represent the expanded and contracted gene families in this node, respectively. The silhouette image of each species was drawn according to their morphology using Adobe Illustrator software. **b**, Relative evolutionary rates of the flatfish species. Zebrafish was used as the outgroup and *Platichthys stellatus* as the reference species. The arrow represents the trend of the relative evolutionary rates. The colored ovals represent different fish groups that showed contrasted relative evolutionary rates.

we calculated the relative evolutionary rates in RFP, FLP and closely related Perciformes species using single-copy orthologous genes. Our results revealed a much higher relative evolutionary rate in RFP than in Perciformes species (Fig. 2b, Supplementary Tables 81 and 82 and Supplementary Note 18). The relative evolutionary rate of FLP is also slightly higher than that for Perciformes species (Fig. 2b and Supplementary Tables 81 and 82), which may explain why they exhibit a 'simply an asymmetric percoid' phenotype compared to RFP. The higher relative evolutionary rates in both RFP and FLP indicate the possible selection pressure that they experienced, although other factors, such as limited population size and rapid drift, could not be excluded[38].

Genes undergoing significant alterations in flatfishes. The fast evolution in both RFP and FLP may predict marked changes in their genomes, which had facilitated evolution of their new body plan after seafloor colonization. To test this hypothesis, we performed comprehensive comparisons on genomic elements among RFP, FLP and other nonflatfish outgroup species (*Larimichthys crocea, Labrus bergylta, Oreochromis niloticus, Oryzias latipes* and *Danio rerio*). We first analyzed gene families that changed rapidly in gene number during the evolution process (see Supplementary Note 19), and identified some expanded and contracted gene families ($P < 0.05$) in RFP and FLP, respectively (Supplementary Tables 82–89 and Supplementary Note 19). However, none of these gene families includes those currently known to mediate body plan development (Extended Data Fig. 2 and Supplementary Note 19), such as WNT, RA, BMP, FGF, NOTCH and HOX[39,40], suggesting involvement of other molecular mechanisms for the unique body plan formation in RFP and FLP. Therefore, we further identified genes undergoing positive selection (PSGs) or rapid evolution (REGs) or containing lineage-specific mutations (LSGs) (Supplementary Figs. 18–23 and Supplementary Notes 20 and 21) or lineage-specific conserved noncoding elements (SCNEs) in RFP and FLP (see Supplementary Note 22), respectively.

The enrichment categories of top candidate genes under significant alteration in both RFP and FLP are associated with visual perception (*dmbx1a* (ref. [41]) and *opn3* (ref. [42]) in RFP versus *cryba4* (ref. [43]) and *opn3* (ref. [42]) in FLP), immune response (*bahd1* (ref. [44]), *ripk1* (ref. [45]) and *pik3ip1* (ref. [46]) in RFP versus *nfkbid* (ref. [47]), *trim59* (ref. [48]) and *themis2* (ref. [49]) in FLP), hypoxia tolerance (*fbxl5* (ref. [50]) in RFP versus *ucp2* (ref. [51]) in FLP) and cardiac function (*tmem43* (ref. [52]), *dis3l1* (ref. [53]), *popdc2* (ref. [54]) and *glrx1* (ref. [55]) in RFP versus *irx4a* (ref. [56]) and *glrx3* (*ref.* [57]) in FLP) (Supplementary Tables 90–97, Extended Data Fig. 3 and Supplementary Notes 20 and 21), possibly suggesting a similar remodeling of their visual, immune, respiratory and circulatory systems in benthic adaptation to seafloor colonization (Extended Data Fig. 3 and Supplementary Note 21). Among them, cardiovascular adaptation is a surprising association with rapid sequence evolution during transition from the water column to benthic colonization. Our results revealed that this process may involve not only a cardiac morphological reorganization resulting from selective pressure on cardiac morphogenesis genes (*popdc2* and *irx4a*)[54,56], but also cardiac functional remodeling resulting from selective pressure on genes associated with cardiac conducting efficiency (*tmem43* and *popdc2*)[52,54] and antioxidant capacities (*glrx1* and *glrx3*)[55,57] (Extended Data Fig. 3, Supplementary Tables 90–97 and Supplementary Notes 20, 21 and 23). Such structural and functional alterations of the cardiovascular system in both RFP and FLP might have contributed to their reinforced cardiac output, which is the highest known among teleosts[58,59], and the enhanced antioxidant capacity to cope with hypoxia readily encountered during burrowing into the substrate[58]. Enriched categories of the remaining top candidate genes under significant alterations in RFP and FLP were associated with axial patterning, neural patterning, musculoskeletal restructuring, lipid deposition and fin cartilage reorganization (Supplementary Tables 90–99 and Supplementary Notes 20–22), suggesting their roles in new body plan evolution and adaptation after seafloor colonization.
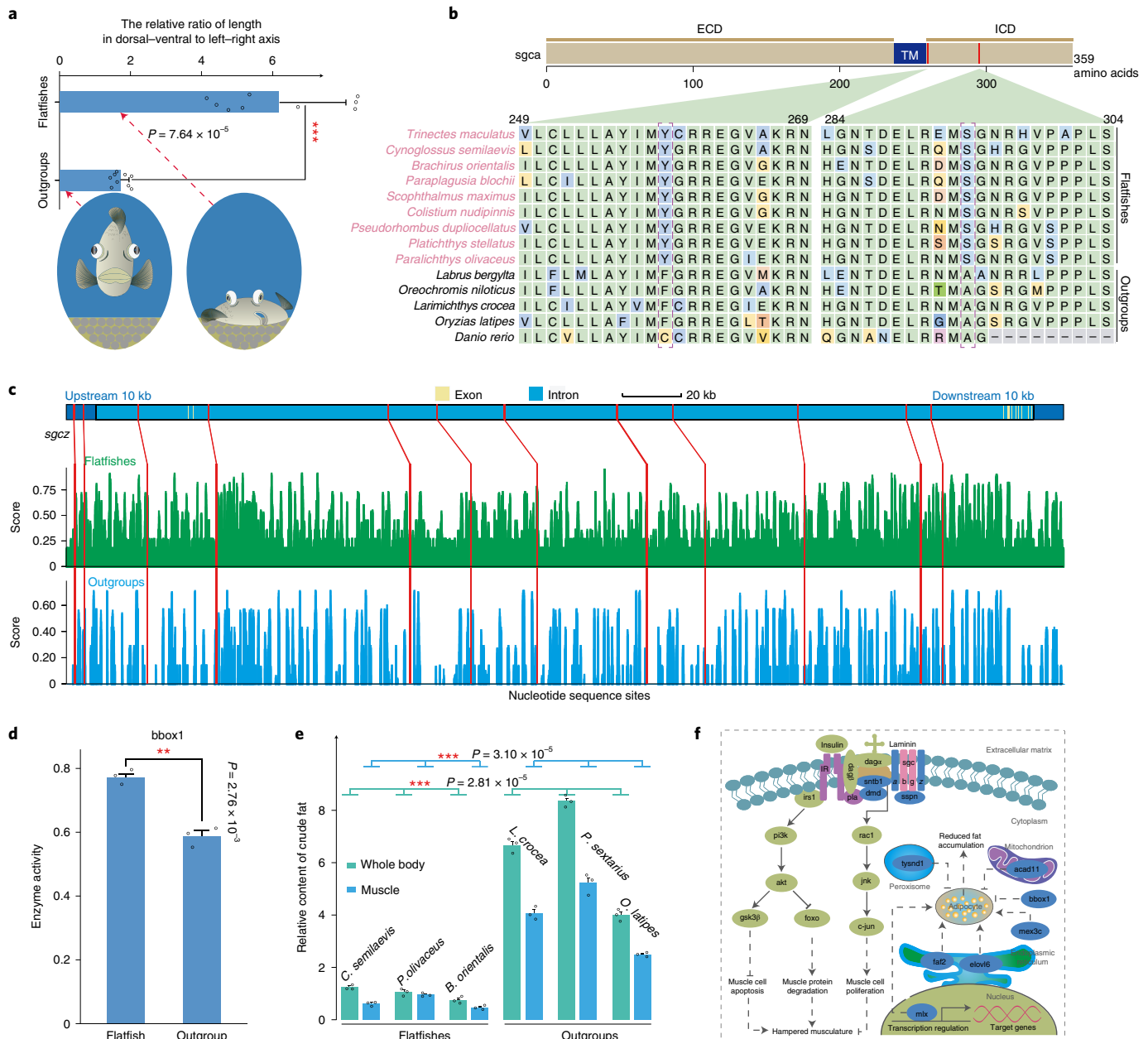
Fig. 3 | Genetic changes correlated with the flat body plan in flatfishes. a, The relative ratio of length in dorsal–ventral to left–right axis. The relative ratio of maximum height of dorsal–ventral axis/maximum length of left–right axis here was used to indicate the degree of body plan flatness of fishes. The ratio was measured in three individuals for each species, and the data are presented as mean values ±s.d. The statistical difference between groups was calculated using Student's *t*-test (two tails) with *** representing a statistical *P* value <0.001. **b**, Mutations in RFP *sgca* compared to outgroups. ECD, ICD and TM represent the extracellular, intracellular and transmembrane domain, respectively. The fixed substitutions between RFP and outgroups are marked with a dashed box. **c**, SCNEs nearby *sgcz* gene. The x axis represents the nucleotide sequence sites across the *sgcz* gene and the y axis represents sequence similarity scores. The green and blue columns represent the average sequence similarity score within RFP and outgroups for each site, respectively; the red lines represent the physical locations of the SCNEs across the gene. **d**, The relative catalyzing activity of bbox1 in RFP and outgroups. The experiment was carried out three times and the data are presented as indicated above. **e**, Relative fat content in whole body and muscle tissues in flatfishes and outgroups. The fat content was measured in three individuals for each species, and the data are presented as indicated above. **f**, Hypothetical signaling pathway that may correlate with the body-plan flatness of flatfishes. Proteins marked blue are those encoded by genes that have undergone genetic alterations in RFP.

Genetic changes correlated with the flat body in flatfishes. The observed enrichment of genes associated with musculoskeletal restructuring and lipid deposition may reflect their roles in the evolution of body plan flatness after metamorphosis in flatfishes (Fig. 3a, Extended Data Fig. 4, Supplementary Fig. 24 and Supplementary Note 24). Such a phenotype possibly confers a selective advantage on the seafloor, where flatfishes usually hide from their enemies

by embedding themselves into a thin layer of substrate, with only the eyes exposed[3,60]. Our comparative genomic analyses, using *Larimichthys crocea*, *Labrus bergylta*, *Oreochromis niloticus*, *Oryzias latipes* and *Danio rerio* as the outgroups, revealed that four genes associated with musculature development have undergone marked alteration in RFP, including the sarcolemma gene *sspn* (PSGs, $P = 8.43 \times 10^{-3}$), the sarcoglycan genes *sgca* (PSGs, $P = 3.30 \times 10^{-3}$)

and *sgcz* (SCNEs), and the dystrophin gene *dmd* (SCNEs) (Fig. 3b,c, Supplementary Tables 90, 91, 98 and 99 and Supplementary Notes 20 and 22). Unexpectedly, all four of these genes are the core components of the dystrophin–glycoprotein complex (DGC), critical in both mechanical stabilization[61] and signal-dependent-activated development of muscular tissues[62–64]. Mutations or abnormal expression of these four genes could cause severe muscular dystrophy or substantial reduction in muscle size in vertebrates[65–68], including zebrafish[69,70]. Among these genes, *sgca* has been the most frequently reported locus that causes the majority of sarcoglycanopathies (one of the severe muscular dystrophies) in humans[71]. Our analysis revealed two RFP-specific missense substitutions in *sgca*, compared to nonflatfish outgroups (Fig. 3b). Both mutations locate within a conserved C-terminal intracellular domain (Fig. 3b), which is critical in the signal-dependent-activated development of muscular tissues. Mutations of this domain frequently cause hampered musculature development and severe muscular dystrophy, such as limb-girdle muscular dystrophy syndrome in humans[72,73]. Such alterations in *sgca* may change the signal-dependent-activation process of muscular development in RFP and thus may have implications in their thinner musculature and flat phenotype.

In addition to musculature, three genes related to lipid metabolism, including *bbox1* (REGs, $P=5.01\times10^{-4}$), *mex3c* (REGs, $P=4.20\times10^{-2}$) and *mlx* (REGs, $P=2.13\times10^{-2}$) also underwent marked changes in RFP (Supplementary Tables 90 and 91 and Supplementary Note 20). These three genes encode either enzymes[74] or metabolic signals[75,76] essential for adipogenesis and fat accumulation in vertebrates. Mutations or abnormal expressions of *mex3c* and *mlx* can result in reduced adiposity and lean phenotype in both mouse[75] and fruit fly[76]. Genetic disruption of *bbox1* leads to a modified serum carnitine level and less fat accumulation in mice[74]. Our in vitro enzyme catalytic activity assay further shows that the RFP-specific bbox1 has significantly higher ($P=2.76\times10^{-3}$) catalytic activity transforming γ-butyrobetaine into L-carnitine (Fig. 3d, Supplementary Fig. 18 and Supplementary Note 25), a molecule critical for fat oxidization and hence fat accumulation[74]. The increased catalytic activity of bbox1 in RFP indicates fast lipid oxidization and decreased fat accumulation in RFP and thus may correlate with the flat phenotype, as observed in other teleosts[77,78]. This hypothesis was further supported by the lower whole body and muscular fat content in RFP compared with other nonflatfish teleosts, as observed in our analysis and in multiple other studies[79–81], since a significantly lower fat content ($P<0.001$) was observed in RFP than in nonflatfish outgroups in both whole body (6.22-fold low) and muscular (5.76-fold low) tissues (Fig. 3e; see Supplementary Note 24). Similar to what was observed in RFP, we also found the DGC component gene (*sntb1*)[82] and lipogenesis-related genes (*bbox1*, *mex3c*, *faf2*, *acad11*, *elovl6* and *tysnd1*)[83–86] to be rapidly evolving in FLP, particularly in *bbox1* (REGs, $P=8.75\times10^{-6}$) and *mex3c* (REGs, $P=1.20\times10^{-5}$) (Supplementary Table 91 and Supplementary Note 20). These observations suggest that a similar mechanism might have been involved in the evolution of flat body plan in both RFP

and FLP. Taken together, our analyses provide evidence that marked changes in musculature development and lipid accumulation genes have occurred in flatfishes, and thus may correlate with the evolutionary origin of their body flatness (Fig. 3f).
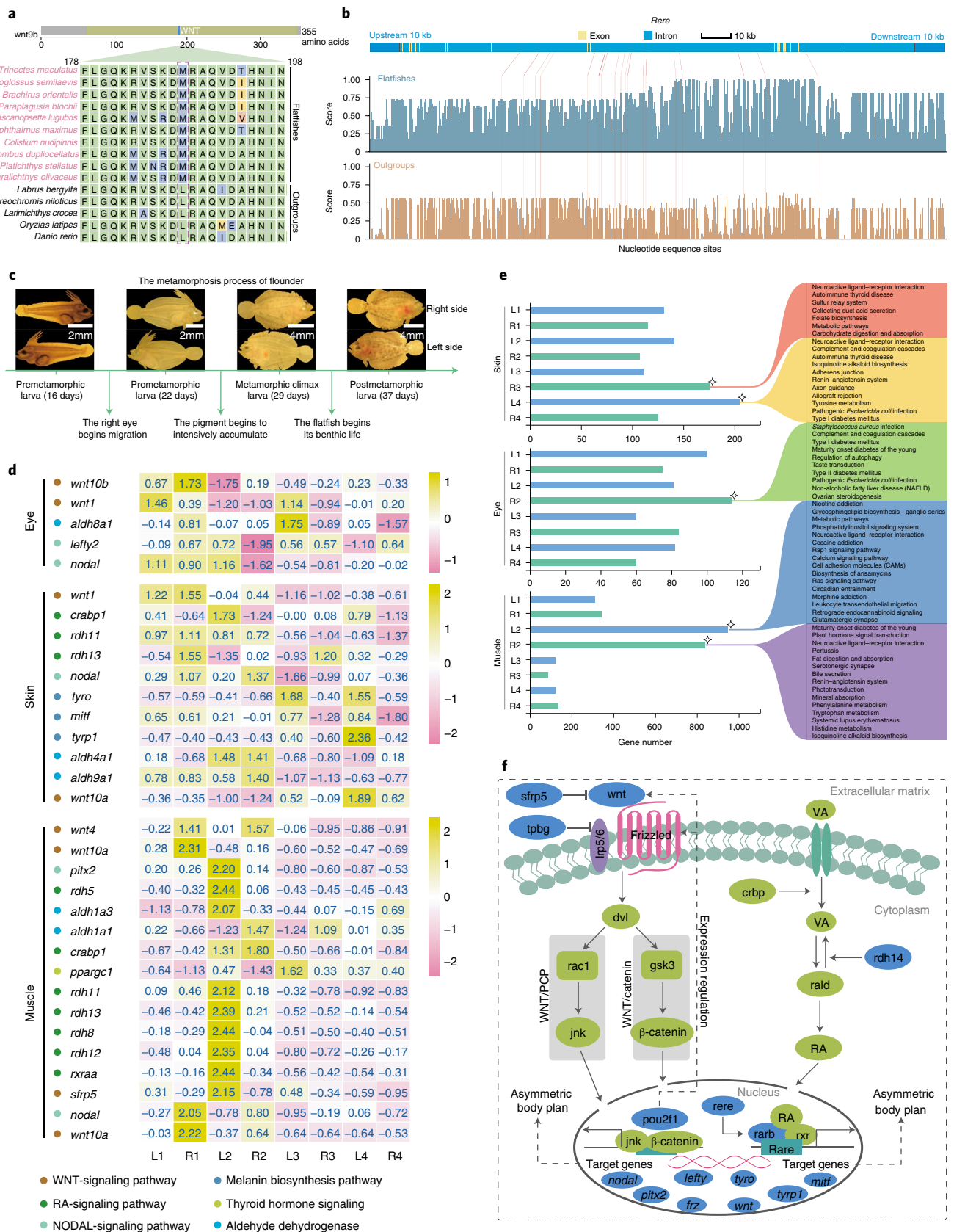
Genetic changes correlate with asymmetric body plan in flatfishes. The body asymmetry is another striking feature of flatfishes, yet its genetic basis remains largely unknown since the time of Darwin[8]. Recently, advances have been made in understanding the genetic regulation of body asymmetry in animals. Such regulation involves several gene families and signal pathways, such as RA, WNT and NODAL[19,87,88]. Both NODAL and RA signals have also been implicated in the body plan asymmetry of flatfishes[16–19]. Our comparative genomic analyses showed that multiple genes from WNT and RA signal pathways have undergone remarkable genetic alterations in RFP (see Supplementary Notes 20–22), suggesting their roles in the evolution of asymmetric body plan. These WNT-signaling genes include *wnt9b* (LSGs, L188M), *sfrp5* (LSGs, K236R), *tpbg* (PSGs, $P=8.02\times10^{-4}$), *pou2f1* (REGs, $P=2.94\times10^{-3}$) (Fig. 4a, Supplementary Tables 90 and 96 and Supplementary Notes 20 and 21), which encode either ligands (for example, wnt9b) or direct modulators (for example, pou2f1 (ref. [89]), tpbg (ref. [90]) and sfrp5 (ref. [91])) of the WNT-signaling pathway. Defects or expression disruptions of *wnt9b*, *tpbg*, *sfrp5* and *pou2f1* genes lead to deficiency in the WNT signal pathway, and bilateral craniofacial asymmetry and skull malformation in vertebrates[92,93], including zebrafish[94]. Our analyses also revealed substantial changes in physicochemical properties and three-dimensional structure of these WNT components in RFP (for example, T212P and P428S cause polarity changes of amino acids in pou2f1; T169K caused charge changes in tpbg; K236R caused protein structure changes in sfrp5) (Supplementary Figs. 19–22). The alterations of so many axial-patterning WNT-signaling pathway genes may indicate their role in the body plan asymmetry of RFP. Similarly, three RA-signaling pathway genes have also undergone significant alteration in RFP, that is, *rdh14* (REGs, $P=3.69\times10^{-3}$), *rere* (SCNEs) and *rarb* (SCNEs) (Fig. 4b, Supplementary Tables 90, 98 and 99 and Supplementary Notes 20 and 22). These genes encode core components in the RA signal pathway[95–97] and defects in *rdh*, *rere* or *rar* genes were observed to cause RA-signaling alteration, which results in multiple congenital abnormalities, including bilateral asymmetry of eyes, craniums or somites in vertebrates[97,98]. Our enzyme catalytic activity assay further lends support for such functional alterations in these RA-signaling genes. Compared with the outgroups, RFP-specific rdh14 has much lower (2.51-fold low; $P=2.84\times10^{-6}$) activity catalyzing retinaldehyde into retinol (Extended Data Fig. 5 and Supplementary Note 25), implying more retinaldehyde (substrate for RA synthesis) accumulation and thus RA signal alterations in RFP. These RFP-specific mutations in the RA-signaling genes might have played roles in the asymmetric body plan of RFP, although their actual role still awaits further verification. Interestingly, we noted only two WNT signal pathway genes, *wnt4a* (REGs, $P=0.00$) and *tpbg* (REGs, $P=4.68\times10^{-2}$),

Fig. 4 | Genetic changes associated with asymmetric body plan in flatfishes. a, Mutations of *wnt9b* in RFP compared with outgroups. The diagram of protein structure is shown on the top of the graph, and the site that showed variation is marked with a dashed box. **b**, SCNEs nearby the *rere* gene. The *x* axis represents the nucleotide sequence sites across *rere*; the *y* axis represents sequence similarity scores. The blue and brown columns represent the average sequence similarity score within RFP and outgroups for each site, respectively; the red lines represent the physical locations of the SCNEs across the gene. **c**, Metamorphic process of flounder. Images of flatfish in premetamorphic stage, prometamorphic stage, metamorphic climax stage and postmetamorphic stage are shown. **d**, Left–right asymmetrical expression of genes in eye, skin and muscle tissues of metamorphic larvae of flounder. The numbers 1–4 on the *x* axis represent the four metamorphic stages of flounder. L and R on the *x* axis represent tissues on the left side and right side of the larva, respectively. **e**, Number of the specifically highly expressed genes during metamorphosis of flounder. The right diagram represents the biological processes in which these genes were involved. The rhombus symbols mark the stages during which the number of the genes began to show greatest expression changes. **f**, Hypothetical signaling pathway that may correlate to the body plan asymmetry in flatfishes. Proteins marked blue are those encoded by genes that experienced genetic alteration in RFP. Genes marked blue and italicized are regulating targets of these proteins and also exhibited abnormal expression (left–right asymmetry) during metamorphosis in RFP.

undergoing rapid evolution in FLP (Supplementary Table 91 and Supplementary Note 20). It remains to be elucidated whether such a distinction is related to the less extensive cranial asymmetry usu-

ally observed in FLP compared to typical RFP[14]. However, such a distinction between RFP and FLP provides further evidence for the polyphyletic origin of flatfishes.

Fig. (a) wnt9b amino acid alignment across flatfishes and outgroups. (b) Rere gene region score plots for flatfishes and outgroups across nucleotide sequence sites. (c) The metamorphosis process of flounder, showing premetamorphic larva (16 days), prometamorphic larva (22 days), metamorphic climax larva (29 days) and postmetamorphic larva (37 days); right side and left side views. (d) Heatmaps of gene expression (L1–R4) in Eye, Skin and Muscle for WNT-signaling pathway, RA-signaling pathway, NODAL-signaling pathway, Melanin biosynthesis pathway, Thyroid hormone signaling and Aldehyde dehydrogenase. (e) Gene number bar plots and pathway enrichment for Skin, Eye and Muscle. (f) Schematic of WNT/PCP, WNT/catenin and RA signaling pathways regulating asymmetric body plan.

Our transcriptomic analyses lend further support to the involvement of WNT and RA signaling in the body plan asymmetry of flatfishes. Using *Paralichthys olivaceus* as a representative example, we show that multiple genes in both RA- (*aldh1, aldh8, rdh5, rdh7, rdh8, rdh11, rdh12, rdh13*) and WNT- (*wnt1, wnt4, wnt10*) signaling pathways exhibited obvious transient expression fluctuations in all three examined flounder tissues (eye, muscle, skin) during metamorphosis, with marked left–right asymmetrical expression (both in gene expression level and in specific highly expressed gene number) initiating from the premetamorphic stage, climbing to an asymmetrical climax during the prometamorphic and metamorphic climax stage and then recovering to symmetry in the postmetamorphic stage (Fig. 4c–e, Extended Data Fig. 6, Supplementary Figs. 25 and 26, Supplementary Tables 100–124 and Supplementary Notes 26 and 27). Such gene expression asymmetry and fluctuations during metamorphosis were further confirmed by our real-time quantitative PCR analysis (Extended Data Fig. 7 and Supplementary Note 28). The conspicuous asymmetrical expression of these genes observed during metamorphosis (Fig. 4d,e, Extended Data Figs. 6 and 7 and Supplementary Notes 27 and 28) indicating gradients of WNT and RA signals across the left–right axis, may be related to eye migration, cranium deformation and lopsided pigmentation during metamorphosis. This is again supported by the evidence that the left deviation of expression of pigmentation genes, such as *tyro*[99], *mitf*[99] and *tyrp1* (ref. [99]), usually occurs after the asymmetrical expression of RA and WNT signals in the skin of metamorphosing flounder larvae (Fig. 4c,d, Extended Data Fig. 7 and Supplementary Notes 27 and 28). Left–right asymmetric expression of NODAL-signaling genes (including *nodal, lefty* and *pitx2*) was also observed in the tissues of metamorphic flounder larvae (for example, muscles and eyes) (Fig. 4d). Such obvious reactivation of NODAL signaling in metamorphosis, which is not usually observed in teleosts with a regular body plan[18], is believed to have initiated the left–right asymmetry of flatfishes[16–18]. Yet it remains an open question as to whether such reactivation of NODAL signals can also be attributed to the asymmetrical RA and WNT signals, although cross-talk between them has long been documented in diverse taxa[100,101]. Taken together, our analyses provide gene evolution and expression evidence for the possible involvement of WNT combined with RA-signaling pathways in shaping the asymmetric body plan in flatfishes (Fig. 4f), although the exact role of these RA and WNT genes in the body plan asymmetry still awaits further investigation.

Genetic changes associated with the modified fins and over-substrate maneuvering of flatfishes. To fit their specialized body plan, flatfishes have also evolved a new vertebrate gait of 'fin-feet' walking, which enables their flexible horizontal maneuvering over substrates[5,6] while keeping both of their eyes alert from above[60]. Such fin-feet walking was attributed to their largely elongated median fins[11] and their largely reduced paired fins (for example, pectoral fins[11]) (Extended Data Figs. 8 and 9, Supplementary Fig. 27 and Supplementary Note 29), since these specialized fins enable a repeated generation of the 'fin-feet' (mainly by dorsal and anal fins) pushing down against the substrate to produce constant forward movement while keeping an accurate maneuvering orientation (mainly by pectoral fins)[5,6]. However, the genetic basis of such fin phenotypes in flatfishes is unknown. Our comparative genomic analyses revealed two genes, including *hoxd12a* (K105R) and *bhlha9* (PSGs, $P = 9.38 \times 10^{-3}$), underwent considerable changes in RFP (Extended Data Fig. 10, Supplementary Tables 90 and 97 and Supplementary Notes 20 and 21). Among them, *hoxd12a* is closely associated with fin patterning and morphogenesis in teleosts[102], since it encodes a DNA-binding transcription factor essential for regulation of the anterior–posterior pattern of fins[102]. The alterations in *hoxd12a* were believed to account for the forelimb (homologs of teleost fins) reorganization in cetaceans[103] and paired fin degeneration in lung-

fishes[104]. Furthermore, *hoxd12a* has also been implicated in dorsal fin development in flounders[105]. The observed mutations in *hoxd12a* may have implications for the morphological changes of median and paired fins in RFP (Supplementary Table 97 and Supplementary Note 21), although the causative effect of these mutations still awaits further verification. Similarly, *bhlha9* encodes a transcription factor closely related to fin morphogenesis. Knockdown of *bhlha9* in zebrafish usually results in a size reduction of pectoral fins[106]. The observed positive selection in *bhlha9* gene of RFP may also indicate its possible role in fin modification (Supplementary Table 90 and Supplementary Note 20). The *bhlha9* gene is also rapidly evolving ($P = 2.03 \times 10^{-4}$) in FLP (Supplementary Table 91 and Supplementary Note 20), and *hoxd12a* experienced convergent variation between FLP and RFP (Supplementary Table 97 and Supplementary Note 21), indicating a possible role of these two genes in shaping the specialized fin morphologies of RFP and FLP.

Discussion

Our study demonstrates the strengths of combining phylogenomics and comparative genomics to shed light on the evolutionary history and mechanisms of a nonmodel taxon with complex adaptive traits, such as flatfishes. Using large-scale genomic data, we revealed a polyphyletic origin of flatfishes, with real flatfish Pleuronectoidei and flatfish-like Psettodoidei independently evolving from their different percoid ancestors. However, Pleuronectoidei and Psettodoidei also share convergent alterations in genes related to muscular development, lipid accumulation, body axis determination and fin pattern regulation. Meanwhile, Psettodoidei also exhibited unique mutations that may contribute to their less asymmetric body plan compared to Pleuronectoidei. The results obtained in this study have substantially clarified the long-standing controversies over the phylogeny of flatfishes, while the genes highlighted in this study lay a blueprint for future functional characterization of the molecular mechanisms underlying the unusual body plan of flatfishes. The genetic basis of such complex traits in flatfishes will not only enrich our knowledge on how the symmetric body plan that dominates the animal kingdom has evolved, been retained and modified, but also potentially help to unveil congenital causes of similar human pathological disorders, such as muscular atrophy and craniofacial malformations.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00836-9.

References

1. Dornbos, S. Q., Bottjer, D. J. & Chen, J. Y. Paleoecology of benthic metazoans in the early cambrian maotianshan shale biota and the middle cambrian burgess shale biota: evidence for the Cambrian substrate revolution. *Palaeogeogr. Palaeoclmatol. Palaeoecol.* **220**, 47–67 (2005).
2. Bottjer, D. J. The Cambrian substrate revolution and early evolution of the phyla. *J. Earth Sci.* **21**, 21–24 (2010).
3. Ryer, C. H. A review of flatfish behavior relative to trawls. *Fish. Res.* **90**, 138–146 (2008).
4. Janvier, P. Squint of the fossil flatfish. *Nature* **454**, 169–170 (2008).
5. Holmes, R. A. & Gibson, R. N. A comparison of predatory behavior in flatfish. *Anim. Behav.* **31**, 1244–1255 (1983).
6. Fox, C. H., Gibb, A. C., Summers, A. P. & Bemis, W. E. Benthic walking, bounding, and maneuvering in flatfishes (Pleuronectiformes: Pleuronectidae): new vertebrate gaits. *Zoology* **130**, 19–29 (2018).
7. Mivart, St, G. *On the Genesis of Species* (MacMillian, 1871).

8. Darwin, C. *The Origin of Species by Means of Natural Selection* 6th edn (John Murray, 1872).

9. Campbell, M. A., Chen, W. J. & Lopez, J. A. Are flatfishes (Pleuronectiformes) monophyletic? *Mol. Phylogenet. Evol.* **69**, 664–673 (2013).

10. Chapleau, F. Pleuronectiform relationships: a cladistic reassessment. *Bull. Mar. Sci.* **52**, 516–540 (1993).

11. Gudger, E. W. Abnormalities in flatfishes (heterosomata) I. Reversal of sides a comparative study of the known data. *J. Morphol.* **58**, 1–39 (1935).

12. Shi, W. et al. Flatfish monophyly refereed by the relationship of *Psettodes* in Carangimorphariae. *BMC Genomics* **19**, 400 (2018).

13. Betancur-R, R. et al. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Syst. Biol.* **62**, 763–785 (2013).

14. Campbell, M. A., Chen, W. J. & Lopez, J. A. Molecular data do not provide unambiguous support for the monophyly of flatfishes (Pleuronectiformes): a reply to Betancur-R and Orti. *Mol. Phylogenet. Evol.* **75**, 149–153 (2014).

15. Inui, Y. & Miwa, S. Thyroid-hormone induces metamorphosis of flounder larvae. *Gen. Comp. Endocrinol.* **60**, 450–454 (1985).

16. Hashimoto, H. et al. Isolation and characterization of a Japanese flounder clonal line, reversed, which exhibits reversal of metamorphic left-right asymmetry. *Mech. Dev.* **111**, 17–24 (2002).

17. Hashimoto, H. et al. Embryogenesis and expression profiles of charon and nodal-pathway genes in sinistral (*Paralichthys otivaceus*) and dextral (*Verasper variegatus*) flounders. *Zool. Sci.* **24**, 137–146 (2007).

18. Suzuki, T. et al. Metamorphic *pitx2* expression in the left habenula correlated with lateralization of eye-sidedness in flounder. *Dev. Growth Differ.* **51**, 797–808 (2009).

19. Shao, C. W. et al. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* **49**, 119–124 (2017).

20. Liu, L. A., Yu, L. L. & Edwards, S. V. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302 (2010).

21. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

22. Betancur-R, R. et al. The tree of life and a new classification of bony fishes. *PLoS Curr.* https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288 (2013).

23. Li, C., Betancur-R, R., Smith, W. L. & Ortí, G. Monophyly and interrelationships of Snook and Barramundi (Centropomidae sensu Greenwood) and five new markers for fish phylogenetics. *Mol. Phylogen. Evol.* **60**, 463–471 (2011).

24. Hillis, D. M. & Moritz, C. *Molecular Systematics* (Sinauer Associates, 1990).

25. Li, W. *Molecular Evolution* (Sinauer Associates, 1997).

26. Regan, C. T. The origin and evolution of the teleostean fishes of the order Heterosomata. *Ann. Mag. Nat. Hist.* **6**, 484–496 (1910).

27. Hubbs, C. L. Phylogenetic position of the Citharidae, a family of flatfishes. *Misc. Publ. Mus. Zool. Univ. Mich.* **63**, 1–38 (1945).

28. Amaoka, K. Studies on the sinistral flounder found in the waters around Japan: taxonomy, anatomy, and phylogeny. *J. Shimonoseki Univ. Fish.* **18**, 65–340 (1969).

29. Chabanaud, P. Les téléostéens dyssymétriques du Mokatam Inférieur de Tourah. *Mém. Inst. Egypte* **31**, 1–122 (1937).

30. Nelson, J. *Fishes of the World* (Wiley, 2016).

31. Munroe, T. A. in *Flatfishes: Biology and Exploitation* (ed. Gibson, R. N.) Ch. 2 (Wiley, 2005).

32. Harrington, R. C. et al. Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evol. Biol.* **16**, 224–238 (2016).

33. Priede, I. G. & Froese, R. Colonization of the deep sea by fishes. *J. Fish. Biol.* **83**, 1528–1550 (2013).

34. Near, T. J. et al. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl Acad. Sci. USA* **109**, 13698–13703 (2012).

35. Guinot, G. & Cavin, L. 'Fish' (Actinopterygii and Elasmobranchii) diversification patterns through deep time. *Biol. Rev. Camb. Philos. Soc.* **91**, 950–981 (2016).

36. Tittensor, D. P. et al. Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).

37. Erwin, D. H. Climate as a driver of evolutionary change. *Curr. Biol.* **19**, 575–583 (2009).

38. Kimura, M. & Ohta, T. On the rate of molecular evolution. *J. Mol. Evol.* **1**, 1–17 (1971).

39. Manuel, M. Early evolution of symmetry and polarity in metazoan body plans. *C. R. Biol.* **332**, 184–209 (2009).

40. Levin, M. Left-right asymmetry in embryonic development: a comprehensive review. *Mech. Dev.* **122**, 3–25 (2005).

41. Wong, L., Weadick, C. J., Kuo, C., Chang, B. S. W. & Tropepe, V. Duplicate *dmbx1* genes regulate progenitor cell cycle and differentiation during zebrafish midbrain and retinal development. *BMC Dev. Biol.* **10**, 100 (2010).

42. Rios, M. N., Marchese, N. A. & Guido, M. E. Expression of non-visual opsins Opn3 and Opn5 in the developing inner retinal cells of birds. Light-responses in Müller glial cells. *Front. Cell. Neurosci.* **13**, 376 (2019).

43. Limi, S. et al. Bidirectional analysis of Cryba4-Crybb1 nascent transcription and nuclear accumulation of Crybb3 mRNAs in lens fibers. *Invest. Ophthalmol. Vis. Sci.* **60**, 234–244 (2019).

44. Lebreton, A. et al. A bacterial protein targets the BAHD1 chromatin complex to stimulate type III interferon response. *Science* **331**, 1319–1321 (2011).

45. Muscolino, E. et al. Herpes viruses induce aggregation and selective autophagy of host signalling proteins NEMO and RIPK1 as an immune-evasion mechanism. *Nat. Microbiol.* **5**, 331–342 (2020).

46. Chen, Y. C. et al. Pik3ip1 is a negative immune regulator that inhibits antitumor T-cell immunity. *Clin. Cancer Res.* **25**, 6180–6194 (2019).

47. Arnold, C. N. et al. A forward genetic screen reveals roles for Nfkbid, Zeb1, and Ruvbl2 in humoral immunity. *Proc. Natl Acad. Sci.* **109**, 12286–12293 (2012).

48. Jin, Z. et al. TRIM59 protects mice from sepsis by regulating inflammation and phagocytosis in macrophages. *Front. Immunol.* **11**, 263 (2020).

49. Deobagkar-Lele, M., Anzilotti, C. & Cornall, R. J. Themis2: setting the threshold for B-cell selection. *Cell. Mol. Immunol.* **14**, 643–645 (2017).

50. Taabazuing, C. Y., Hangasky, J. A. & Knapp, M. J. Oxygen sensing strategies in mammals and bacteria. *J. Inorg. Biochem.* **133**, 63–72 (2014).

51. Chen, M. et al. Ursolic acid stimulates UCP2 expression and protects H9c2 cells from hypoxia-reoxygenation injury via p38 signaling. *J. Biosci.* **43**, 857–865 (2018).

52. Siragam, V. et al. TMEM43 mutation p.S358L alters intercalated disc protein expression and reduces conduction velocity in arrhythmogenic right ventricular cardiomyopathy. *PLoS ONE* **9**, e109128 (2014).

53. Lee, J. Y. et al. Genome-based exome sequencing analysis identifies GYG1, DIS3L and DDRGK1 are associated with myocardial infarction in Koreans. *J. Genet.* **96**, 1041–1046 (2017).

54. Kirchmaier, B. C. et al. The Popeye domain containing 2 (*popdc2*) gene in zebrafish is required for heart and skeletal muscle development. *Dev. Biol.* **363**, 438–450 (2012).

55. Li, S. Y. et al. Protective effect and mechanism of glutaredoxin 1 on coronary arteries endothelial cells damage induced by high glucose. *Biomed. Mater. Eng.* **24**, 3897–3903 (2014).

56. Christoffels, V. M. et al. Chamber formation and morphogenesis in the developing mammalian heart. *Dev. Biol.* **223**, 266–278 (2000).

57. Donelson, J. et al. Cardiac-specific ablation of glutaredoxin 3 leads to cardiac hypertrophy and heart failure. *Physiol. Rep.* **7**, e14071 (2019).

58. Mendonça, P. C., Genge, A. G., Deitch, E. J. & Gamperl, A. K. Mechanisms responsible for the enhanced pumping capacity of the in situ winter flounder heart (*Pseudopleuronectes americanus*). *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **293**, 2112–2119 (2007).

59. Joaquim, N., Wagner, G. N. & Gamperl, A. K. Cardiac function and critical swimming speed of the winter flounder (*Pleuronectes americanus*) at two temperatures. *Comp. Biochem. Phys. A* **138**, 277–285 (2004).

60. Gibson, R. N., Stoner, A. W. & Ryer, C. H. *Flatfishes: Biology and Exploitation* 2nd edn (ed. Gibson, R. N.) Ch. 12 (Wiley, 2015).

61. Crosbie, R. H. et al. Membrane targeting and stabilization of sarcospan is mediated by the sarcoglycan subcomplex. *J. Cell Biol.* **145**, 153–165 (1999).

62. Mutlak, Y. E. et al. Novel signaling hub of insulin receptor, dystrophin glycoprotein complex and plakoglobin regulates muscle size. *Nat. Commun.* **11**, 1381 (2020).

63. Kobayashi, J. et al. Molecular regulation of skeletal muscle mass and the contribution of nitric oxide: a review. *FASEB Bioadv.* **1**, 364–374 (2019).

64. Cassano, M. et al. Cellular mechanisms and local progenitor activation to regulate skeletal muscle mass. *J. Muscle Res. Cell. Motil.* **30**, 243–253 (2009).

65. Marshall, J. L. & Crosbie-Watson, R. H. Sarcospan: a small protein with large potential for Duchenne muscular dystrophy. *Skelet. Muscle* **3**, 1 (2013).

66. Guglieri, M. et al. Clinical, molecular, and protein correlations in a large sample of genetically diagnosed Italian limb girdle muscular dystrophy patients. *Hum. Mutat.* **29**, 258–266 (2008).

67. Wasala, N. B. et al. Genomic removal of a therapeutic mini-dystrophin gene from adult mice elicits a Duchenne muscular dystrophy-like phenotype. *Hum. Mol. Genet.* **25**, 2633–2644 (2016).

68. Araishi, K. et al. Loss of the sarcoglycan complex and sarcospan leads to muscular dystrophy in beta-sarcoglycan-deficient mice. *Am. J. Hum. Genet.* **8**, 1589–1598 (1999).

69. Bassett, D. I. et al. Dystrophin is required for the formation of stable muscle attachments in the zebrafish embryo. *Development* **130**, 5851–5860 (2003).

70. Parsons, M. J., Campos, I., Hirst, E. M. & Stemple, D. L. Removal of dystroglycan causes severe muscular dystrophy in zebrafish embryos. *Development* **129**, 3505–3512 (2002).

71. Liang, W. C. et al. Probable high prevalence of limb-girdle muscular dystrophy type 2D in Taiwan. *J. Neurol. Sci.* **362**, 304–308 (2016).

72. Monies, D. et al. A first-line diagnostic assay for limb-girdle muscular dystrophy and other myopathies. *Hum. Genomics* **10**, 32 (2016).

73. Xie, Z. Y. et al. Clinical and genetic spectrum of sarcoglycanopathies in a large cohort of Chinese patients. *Orphanet J. Rare Dis.* **14**, 43 (2019).

74. Zhao, M. et al. TMAVA, a metabolite of intestinal microbes, is increased in plasma from patients with liver steatosis, inhibits γ-butyrobetaine hydroxylase, and exacerbates fatty liver in mice. *Gastroenterology* **158**, 2266–2281 (2020).

75. Jiao, Y. et al. *Mex3c* mutation reduces adiposity and increases energy expenditure. *Mol. Cell. Biol.* **32**, 4350–4362 (2012).

76. Sassu, E. D. et al. Mio/dChREBP coordinately increases fat mass by regulating lipid synthesis and feeding behavior in *Drosophila. Biochem. Biophys. Res. Commun.* **426**, 43–48 (2012).

77. Goetz, F. et al. A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol. Ecol.* **19**, 176–196 (2010).

78. Hansen, M. J. et al. Life history differences between fat and lean morphs of lake charr (*Salvelinus namaycush*) in Great Slave Lake, Northwest Territories, Canada. *Hydrobiologia* **783**, 21–35 (2016).

79. Stansby, M. E. Chemical characteristics of fish caught in the northeast Pacific Ocean. *Mar. Fish. Rev.* **38**, 3–11 (1976).

80. Schloesser, R. W. & Fabrizio, M. C. Condition indices as surrogates of energy density and lipid content in juveniles of three fish species. *Trans. Am. Fish. Soc.* **146**, 1058–1069 (2017).

81. Wander, R. C. & Patton, B. D. Lipids and fatty acids of three species of northeast pacific finfish harvested in summer. *J. Food Compos. Anal.* **4**, 128–135 (1991).

82. Yoshizawa, K. et al. Analyses of beta-1 syntrophin, syndecan 2 and Gem GTPase as candidates for chicken muscular dystrophy. *Exp. Anim.* **52**, 391–396 (2003).

83. Liang, X. J. et al. Transcriptional response of subcutaneous white adipose tissue to acute cold exposure in mice. *Int. J. Mol. Sci.* **20**, 3680 (2019).

84. Matsuzaka, T. & Shimano, H. Elovl6: a new player in fatty acid metabolism and insulin sensitivity. *J. Mol. Med.* **87**, 379–384 (2009).

85. Mizuno, Y. et al. Tysnd1 deficiency in mice interferes with the peroxisomal localization of PTS2 enzymes, causing lipid metabolic abnormalities and male infertility. *PLoS Genet.* **9**, e1003286 (2013).

86. Wang, C. W. & Lee, S. C. The ubiquitin-like (UBX)-domain-containing protein Ubx2/Ubxd8 regulates lipid droplet homeostasis. *J. Cell Sci.* **125**, 2930–2939 (2012).

87. Hamada, H., Meno, C., Watanabe, D. & Saijoh, Y. Establishment of vertebrate left–right asymmetry. *Nature* **3**, 103–113 (2002).

88. Oishi, I., Kawakami, Y., Raya, A., Callol-Massot, C. & Belmonte, J. C. I. Regulation of primary cilia formation and left-right patterning in zebrafish by a noncanonical Wnt signaling mediator, *duboraya. Nat. Genet.* **38**, 1316–1322 (2006).

89. Katoh, M. Comparative genomics on Wnt3-Wnt9b gene cluster. *Int. J. Mol. Med.* **15**, 743–747 (2005).

90. Kagermeier-Schenk, B. et al. Waif1/5T4 inhibits Wnt/β-catenin signaling and activates noncanonical Wnt pathways by modifying LRP6 subcellular localization. *Dev. Cell* **21**, 1129–1143 (2011).

91. Satoh, W., Matsuyama, M., Takemura, H., Aizawa, S. & Shimono, A. Sfrp1, Sfrp2, and Sfrp5 regulate the Wnt/β-catenin and the planar cell polarity pathways during early trunk formation in mouse. *Genesis* **46**, 92–103 (2008).

92. Juriloff, D. M., Harris, M. J., McMahon, A. P., Carroll, T. J. & Lidral, A. C. Wnt9b is the mutated gene involved in multifactorial nonsyndromic cleft lip with or without cleft palate in A/WySn mice, as confirmed by a genetic complementation test. *Birth Defects Res.* **76**, 574–579 (2006).

93. Marini, N. J., Asrani, K., Yang, W., Rine, J. & Shaw, G. M. Accumulation of rare coding variants in genes implicated in risk of human cleft lip with or without cleft palate. *Am. J. Med. Genet.* **179**, 1260–1269 (2019).

94. Jackson, H. W., Prakash, D., Litaker, M., Ferreira, T. & Jezewski, P. A. Zebrafish Wnt9b patterns the first pharyngeal arch into D-I-V domains and promotes anterior-medial outgrowth. *Am. J. Mol. Biol.* **5**, 57–83 (2015).

95. Belyaeva, O. V. & Kedishvili, N. Y. Human pancreas protein 2 (PAN2) has a retinal reductase activity and is ubiquitously expressed in human tissues. *FEBS Lett.* **531**, 489–493 (2002).

96. Lohnes, D., Mark, M., Mendelsohn, C., Dollé, P. & Dierich, A. J. D. Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development* **120**, 2723–2748 (1994).

97. Vilhais-Neto, G. C. et al. Rere controls retinoic acid signalling and somite bilateral symmetry. *Nature* **463**, 953–957 (2010).

98. Vermot, J. et al. Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo. *Science* **308**, 563–566 (2005).

99. Sturm, R. A. & Duffy, D. L. Human pigmentation genes under environmental selection. *Genome Biol.* **13**, 248 (2012).

100. Kioussi, C. et al. Identification of a Wnt/Dvl/β-catenin → Pitx2 pathway mediating cell-type-specific proliferation during development. *Cell* **111**, 673–685 (2002).

101. Wasiak, S. & Lohnes, D. Retinoic acid affects left–right patterning. *Dev. Biol.* **15**, 332–342 (1999).

102. Freitas, R., Zhang, G. J. & Cohn, M. J. Evidence that mechanisms of fin development evolved in the midline of early vertebrates. *Nature* **442**, 1033–1037 (2006).

103. Wang, Z. et al. Adaptive evolution of *5′HoxD* genes in the origin and diversification of the cetacean flipper. *Mol. Biol. Evol.* **26**, 613–622 (2009).

104. Liang, D., Wu, R. G., Geng, J., Wang, C. L. & Zhang, P. A general scenario of *Hox* gene inventory variation among major sarcopterygian lineages. *BMC Evol. Biol.* **11**, 25 (2011).

105. Chen, J., Liu, X. Y., Yao, X. H., Gao, F. & Bao, B. L. Dorsal fin development in flounder, *Paralichthys olivaceus*: bud formation and its cellular origin. *Gene Expr. Patterns* **25**, 22–28 (2017).

106. Klopocki, E. et al. Duplications of BHLHA9 are associated with ectrodactyly and tibia hemimelia inherited in non-Mendelian fashion. *J. Med. Genet.* **49**, 119–125 (2012).

Methods

DNA and RNA extraction. Genomic DNA was isolated from muscle tissues using the classic phenol–chloroform method. Total RNA was extracted using a Trizol kit (Life Technologies). The quality and quantity of extracted DNA/RNA were assessed using an Agilent 2100 bioanalyzer (Agilent Technologies), and their integrity was further evaluated on agarose gel stained with ethidium bromide. The extracted DNA/RNA samples were stored at −80 °C until subsequent library construction and genome/transcriptome sequencing. All tissue sampling and DNA and RNA extraction processes complied with all relevant ethical regulations provided by the Institutional Animals Care and Use Committee of Zhejiang Ocean University and by the Experimental Animal Management and Ethics Committee of South China Sea Institute of Oceanography, Chinese Academy of Sciences.

Library construction and sequencing. For genome sequencing of the seven species of *Trinectes maculatus*, *Chascanopsetta lugubris*, *Brachirus orientalis*, *Paraplagusia blochii*, *Colistium nudipinnis*, *Pseudorhombus dupliocellatus* and *Platichthys stellatus*, both the short-insert (350–700 bp) and long-insert (>1 kb) paired-end libraries of each species were constructed from the extracted genomic DNA of each species using the Illumina library construction kit (NEBNext Ultra DNA Library Prep Kit from Illumina, catalog no. E7370S) and sequenced on the Illumina HiSeq 4000 platform. For genome sequencing of the three species of *Psettodes erumei*, *Toxotes chatareus* and *Polydactylus sextarius*, the extracted genomic DNA was size-selected using PippinHT (Sage Science). Then, the Nanopore libraries were constructed and sequenced on PromethION DNA sequencer (Oxford Nanopore Technologies). Genomes of the three species of *Platichthys stellatus*, *Toxotes chatareus* and *Polydactylus sextarius* were further sequenced on the Hi-C platform to obtain chromosome-level genome assemblies. For Hi-C library construction, DNA extracted from each species was fragmented and purified using magnetic beads. Hi-C libraries were sequenced on the Illumina HiSeq 4000 platform with 150-bp paired-end reads. For RNA sequencing of the four species *Platichthys stellatus*, *Toxotes chatareus*, *Polydactylus sextarius* and *Paralichthys olivaceus*, the complementary DNA libraries were constructed from RNA extracted from various tissues, such as eye, liver, muscle and skin, as indicated in Supplementary Table 2 for different analysis purposes according to the manufacturer's instructions (NEBNext Ultra RNA Library Prep Kit from Illumina, catalog no. E7530S) and sequenced on the Illumina HiSeq 4000 platform.

Quality control of sequencing data. For Illumina sequencing reads, all low-quality reads, duplicated reads and adapter sequences were removed using Perl scripts. For Nanopore long reads, mean quality for each read was calculated and only reads longer than 1 kb with mean quality ≥7 were retained. For Hi-C sequencing data, the low-quality reads were further filtered using Hi-C-Pro software (v.3.2)[107] after prefiltering with Perl scripts.

Genome size estimation. Genome size of each species was estimated using the short-insert library reads by the *k*-mer method. The 17-mer was chosen for *k*-mer analysis in this study, and the genome size (*G*) was estimated with the following formula: $G = K_{num}/K_{depth}$, where K_{num} and K_{depth} represent the total number of 17-mers and the peak of depth of the 17-mer, respectively.

Genome assembly and chromosome construction. For the genome assembly, seven species (*Trinectes maculatus*, *Chascanopsetta lugubris*, *Brachirus orientalis*, *Paraplagusia blochii*, *Colistium nudipinnis*, *Pseudorhombus dupliocellatus* and *Platichthys stellatus*) were assembled with Illumina short reads using the Platanus software (v.1.2.4)[108], and all the cleaned short reads were used to fill the gaps of the genome using Gapcloser (v.1.10). Three species (*Psettodes erumei*, *Toxotes chatareus* and *Polydactylus sextarius*) were assembled with Nanopore long reads using WTDBG software (v.1.2.8)[109], and all the cleaned Illumina short-insert reads were aligned to the assembled contigs to conduct error correction. For chromosome construction of three species (*Platichthys stellatus*, *Toxotes chatareus* and *Polydactylus sextarius*), the filtered Hi-C reads were aligned to the assembled genome and then anchored to chromosomes using three-dimensional de novo assembly software (v.170123)[110].

Ancestral chromosome reconstruction. First, the chromosome-level genomic data of *Platichthys stellatus* and *Cynoglossus semilaevis*, in the real flatfish Pleuronectoidei lineage, and *Toxotes chatareus* and *Polydactylus sextarius* (sequenced in this study), leading to the flatfish-like Psettodoidei lineage, were aligned and the genome synteny was analyzed using LAST[111] with the parameters of --k 1 -m 10 --E 0.05. Then, the chromosome variation events within and between lineages were compared using ANGES (v.1.01)[112] to detect the lineage-specific chromosome variation. Finally, contig sequences obtained from Nanopore reads of *Psettodes erumei* were used to check for these lineage-specific chromosome fusion and fission events to further test if flatfish-like Psettodoidei lineage (including *Psettodes erumei*) has different ancestral chromosomes from that of real flatfish Pleuronectoidei.

Genome annotation. Repetitive sequences were identified using different software programs. Transposable elements (TEs) were annotated on both protein and DNA levels. On the protein level, the RepeatProteinMask (RM-BLASTX) was used to search TEs in its protein database. On the DNA level, RepeatModeler software (v.1.0.8) was used to build de novo repeat library and RepeatMasker (v.4.0.6)[113] was then run against the de novo library and repbase (RepBase v.16.02) separately to identify homologous repeats. Protein-coding genes were annotated using three combined approaches, including de novo prediction, homology-based annotation and/or transcripts-based annotation from the repeats-masked genome. For de novo prediction, Augustus (v.3.2.1)[114] and GENSCAN (v.1.0)[115] were used. For homology-based annotation, protein sequences of seven species (*Mus musculus*, *Gallus gallus*, *Callorhinchus milii*, *Takifugu rubripes*, *Lepisosteus oculatus*, *Cynoglossus semilaevis* and *Paralichthys olivaceus*) were downloaded from NCBI and protein sequences of one species (*Danio rerio*) were downloaded from Ensembl. The longest transcript of each gene was selected and any genes with early termination sites were removed. All remaining genes were aligned to the repeat-masked genome for homology-based annotation using tblastn with *e*-value less than 1×10^{-5}. Genewise software (v.2.2.0)[116] was used to identify the longest coding regions and/or highest score in each gene locus to support the presence of a homologous gene. For transcript-based annotation, cleaned RNA-seq reads were assembled into transcripts, and then were aligned against the assembled genome to link spliced alignments. EvidenceModeler (v.1.1.1)[117] was used to integrate the results derived from these methods into the final gene set. Functions of these predicted genes were analyzed using the public protein databases. InterProScan (v.4.8) was used to screen proteins against databases (Pfam, v.27.0; prints, v.42.0; prosite, v.20.97; ProDom, v.2006.1; smart, v.6.2). In addition, the Kyoto Encyclopedia of Genes and Genomes (KEGG), NR, SwissProt (v.2011.6) and TrEMBL (v.2011.6) databases were also searched for homology-based function assignments using BLAST software (v.2.6.0) with *e*-value of 1×10^{-5}.

Identification of orthologous genes. Orthologs were identified in the assembled genomes of ten sequenced species, along with the species with published genome sequences (*Cynoglossus semilaevis*, *Paralichthys olivaceus*, *Scophthalmus maximus*, *Danio rerio*, *Larimichthys crocea*, *Labrus bergylta*, *Oreochromis niloticus* and *Oryzias latipes*) using the OrthoMCL pipeline (v.2.0.9)[118]. Briefly, all the protein-coding genes of the published species were downloaded from the NCBI database, except for *Scophthalmus maximus*, which was downloaded from its own website (http://denovo.cnag.cat/genomes/turbot). To improve the accuracy of the analysis, genes that encode shorter than 30 amino acids or have early stop codons in the coding regions were removed. All the remaining genes were aligned and reciprocally compared, and the reciprocal best similarity pairs among species were considered as putative orthologs after further evaluation using MCscan software (v.0.9.13)[119].

Phylogenetic tree construction and divergence time evaluation. All the 1,693 single-copy homologous genes identified among species (*Trinectes maculatus*, *Chascanopsetta lugubris*, *Brachirus orientalis*, *Paraplagusia blochii*, *Colistium nudipinnis*, *Pseudorhombus dupliocellatus*, *Platichthys stellatus*, *Psettodes erumei*, *Polydactylus sextarius*, *Toxotes chatareus*, *Cynoglossus semilaevis*, *Paralichthys olivaceus*, *Scophthalmus maximus*, *Danio rerio*, *Larimichthys crocea*, *Labrus bergylta*, *Oreochromis niloticus* and *Oryzias latipes*) were aligned and concatenated into supergenes for phylogenetic relationship analyses. Maximum likelihood-based phylogenetic analysis was conducted using RAxML (v.8.2.9)[120]. Meanwhile, species trees were also constructed using MPEST (v.2.0)[20] and OrthoFinder (v.2.3.5)[21]. Divergence times of these species were then estimated on the basis of the 4dTV sequences via Bayesian relaxed molecular clock approach using MCMCtree program in the PAML package (v.4.8)[121]. Fossil records downloaded from the TIMETREE website (http://www.timetree.org) were used for calibrating our calculated divergence time.

Estimation of relative evolutionary rates. The relative evolutionary rates of species were calculated using two-cluster analysis and Tajima's relative rate test. Two-cluster analysis was performed to test molecular evolution of multiple sequences in the phylogenetic context. A faster or slower evolutionary rate in a particular taxon was analyzed using *Z*-statistics and tpcv module in the LINTRE program. For Tajima's relative rate test, a higher number of lineage-specific substitutions indicates a much faster evolutionary rate using the chi-squared test. All the single-copy genes were used in these two analyses with zebrafish as the outgroup species.

Estimation of gene family expansion and contraction. Expansion and contraction of gene clusters was determined using the CAFE software (v.3.1)[122]. The phylogenetic tree and divergence time analyzed in the previous steps were used in CAFE to infer changes in gene family sizes using a probabilistic model.

Detection of positive selection. All one-to-one orthologous genes extracted from flatfish species and outgroup species (*Larimichthys crocea*, *Labrus bergylta*, *Oreochromis niloticus*, *Oryzias latipes* and *Danio rerio*) were used to identify positively selected or rapidly evolving genes. The multiple sequence alignments were generated and used to estimate three types of ω (the ratio of the rate of nonsynonymous substitutions to the rate of synonymous substitutions) using

branch model in the codeml program of the PAML package (v.4.8)[121]. Branch model (model = 2, NSsites = 0) was used to detect ω of appointed branch to test ($\omega0$) and average ω of all the other branches ($\omega1$) and the mean of whole branches ($\omega2$). Then χ^2 test was used to check whether $\omega0$ was significantly higher than $\omega1$ and $\omega2$ under the threshold P value <0.05, which hinted that these genes would be under positive selection or fast evolution.

Identification of genes with lineage-specific mutation. The high-quality alignments were also used to identify the lineage-specific mutated genes. In this analysis, all single-copy genes among species were checked and any genes with the same variation across all particular taxa, compared with outgroup species, were identified as LSGs. Candidate LSGs were further double-checked using original Illumina reads to avoid assembly and sequencing errors. In addition, Bayesian ancestral state inference conducted using the codeml program in PAML software (v.4.8)[121] was further used to validate the candidate LSGs. In the Bayesian framework, the ancestral state was inferred by the state with the highest posterior probability. In our case, only the ancestral state of Pleuronectoidei was different from the ancestor of all the Pleuronectiformes species, *Toxotes chatareus* and *Polydactylus sextarius*; the potential LSGs were therefore recognized as the true Pleuronectoidei LSGs.

Identification of conserved noncoding elements. Using the *Platichthys stellatus* genome as the reference, the genomes of flatfish and outgroup species were aligned to the reference genome using LAST software (v.802)[111] with the following parameters: -P 5 -m 100 -E 0.05.

The generated alignments were checked locus by locus, and the loci that were present in more than eight Pleuronectoidei species, but absent in any nonflatfish species, were recognized as the potential Pleuronectoidei-specific conserved noncoding elements. Any SCNE sequences less than 20 bp were removed to ensure the accuracy of identification.

Gene expression profile analysis. RNA extracted from eye, skin and muscle tissues across the left–right axis in different metamorphic time windows (premetamorphic larva, prometamorphic larva, metamorphic climax larva and postmetamorphic larva) of *Paralichthys olivaceus* was sequenced on the Illumina sequencing platform. For each metamorphic time window, three biological replicates were sampled, with each replicate containing tissues from at least 30 individuals because of the small size of the larvae, and was used for the RNA extraction and sequencing. Raw reads were filtered and remaining high-quality reads were aligned to the assembled genome using Tophat2 (v.2.1.1)[123]. The transcripts were assembled and gene expression values were analyzed using the cufflinks software (v.2.2.1)[124].

Real-time quantitative PCR assay. Real-time quantitative PCR (qPCR) was used to verify the differentially expressed genes across the left–right body axis of *Paralichthys olivaceus*. Samples were collected as indicated above and extracted RNA was used for obtaining cDNA using the PrimeScript RT reagent kit with gDNA Eraser (Perfect Real Time) (TaKaRa, catalog no. RR047A). The qPCR analysis was performed using the TaKaRa TB Green Premix Ex TaqII (Tli RNaseH Plus) reagents (TaKaRa, catalog no. RR820A). The β-actin gene was used as the internal control of the qPCR experiment. Each experiment was performed with three reaction replicates and calculated relative expression value of genes using the detected threshold cycle (Ct) value.

Catalytic activity assay of enzymes. In vitro enzymatic activity assay was used to test the functional consequence of RFP-specific mutation in bbox1 and rdh14 proteins. RFP-specific genes and those of the outgroups were codon-optimized according to the *Escherichia coli* preference and then synthesized and cloned into the vector pET-28a by Wuhan Gene Create Biological Engineering. The plasmid was transformed into DH5α competent cells for amplification, and then the plasmids were extracted for verification. Finally, the correct plasmids were transformed into BL21 (DE3) to be expressed. The expressed proteins were further extracted, purified and the enzyme activity was measured according to Rattner et al.[125] and Cao et al.[126], respectively. Each experiment was performed with three reaction replicates to determine the mean ± s.d. of the catalytic activity value of the enzymes.

Statistical analysis. Significant differences between the groups were assessed with Student's *t*-test (two tails). The chi-squared test or Fisher's exact test were used in the significant analysis of gene ontology enrichment according to the data feature, and the hypergeometric test was used in KEGG. Multiple comparisons were corrected for false discovery rate. The symbols *, ** and *** represent a statistical significance of P values <0.05, 0.01 and 0.001, respectively.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All the sequencing data were deposited at the NCBI (*Platichthys stellatus*: PRJNA592732; *Trinectes maculatus*: PRJNA592733; *Brachirus orientalis*: PRJNA592734; *Paraplagusia blochii*: PRJNA592738; *Chascanopsetta lugubris*: PRJNA592739; *Colistium nudipinnis*: PRJNA592742; *Pseudorhombus dupliocellatus*: PRJNA592743; *Polydactylus sextarius*: PRJNA592744; *Toxotes chatareus*: PRJNA592745; *Psettodes erumei*: PRJNA592748; *Paralichthys olivaceus*: PRJNA632737). Besides, the source data of the fish photos were deposited at the Figshare database (https://doi.org/10.6084/m9.figshare.13664201.v1). Source data are provided with this paper.

Code availability
All public software used in this study is provided in the accompanying Nature Research Reporting Summary.

References
107. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
108. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
109. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2019).
110. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
111. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
112. Jones, B. R., Rajaraman, A., annier, E. & Chauve, T. C. ANGES: reconstructing ANcestral GEnomeS maps. *Bioinformatics* **28**, 2388–2390 (2012).
113. Bedell, J. A., Korf, I. & Gish, W. MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics* **16**, 1040–1041 (2000).
114. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, 215–225 (2003).
115. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
116. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
117. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, 7 (2008).
118. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
119. Tang et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
120. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
121. Yang, Z. H. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
122. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
123. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, 36 (2013).
124. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
125. Rattner, A., Smallwood, P. M. & Nathans, J. Identification and characterization of all-trans-retinol dehydrogenase from photoreceptor outer segments, the visual cycle enzyme that reduces all-trans-retinal to all-trans-retinol. *J. Biol. Chem.* **275**, 11034–11043 (2000).
126. Cao, Q. et al. Determination of highly soluble L-carnitine in biological samples by reverse phase high performance liquid chromatography with fluorescent derivatization. *Arch. Pharm. Res.* **30**, 1041–1046 (2007).

Author contributions
Y.L., X.K. and W. Wang conceived and supervised the project. X.K., Z. Lü, L.G., H. Luo, Haifeng Jiang, J.Z., W.S. and S.H. collected the samples. Z. Lü, L.G., L.L., B.M., W. Wan, X.C., Z. Li and L.Z. carried out the experiments. Y.R., Y.L., Z.W., Z. Lü, H. Li, Hui Jiang, Y.Z., Y.W., K.W., C.Z. and Y.Q. performed bioinformatic analyses. Y.L., Z. Lü and Y.R.

Extended Data Fig. 1 | Comparison of the annotated coding genes in the species analyzed in this study with those annotated in previously published species. The x-axis shows the length distribution of the mRNA, CDS, exon, and intron sequences in each species, and the y-axis shows the corresponding ratios of each type of sequences in genome for a particular length.

|	BMP	FGF	HOX	NOTCH	WNT	Total	
	17	31	41	15	20	124	*Platichthys stellatus*
	20	30	45	17	22	134	*Pseudorhombus dupliocellatus*
	21	29	28	12	15	105	*Paralichthys olivaceus*
	21	28	41	19	24	133	*Chascanopsetta lugubris*
	25	34	25	17	21	122	*Scophthatmus maximus*
	22	32	45	13	20	132	*Trinectes maculatus*
	21	29	46	14	21	131	*Colistium nudipinnis*
	18	30	43	11	20	122	*Paraplagusia blochii*
	28	34	49	20	24	155	*Cynoglossus semilaevis*
	22	31	44	14	19	130	*Brachirus orientalis*
	25	32	43	14	19	133	*Psettodes erumei*
	19	29	41	13	19	121	*Toxotes chatareus*
	22	29	35	13	22	121	*Polydactylus sextarius*
	26	37	31	17	26	137	*Larimichthys crocea*
	26	41	48	14	23	152	*Labrus bergylta*
	18	24	32	9	14	97	*Oreochromis niloticus*
	26	29	40	15	23	133	*Oryzias latipes*
	24	34	49	19	26	152	*Danio rerio*

Extended Data Fig. 2 | Statistics of gene number in some gene families known to mediate body plan development in each species. The gene number of each specific gene family is shown in blue circle and total gene number known to mediate body plan development is shown in purple circle. The circle sizes are equivalent to the gene number that was observed.
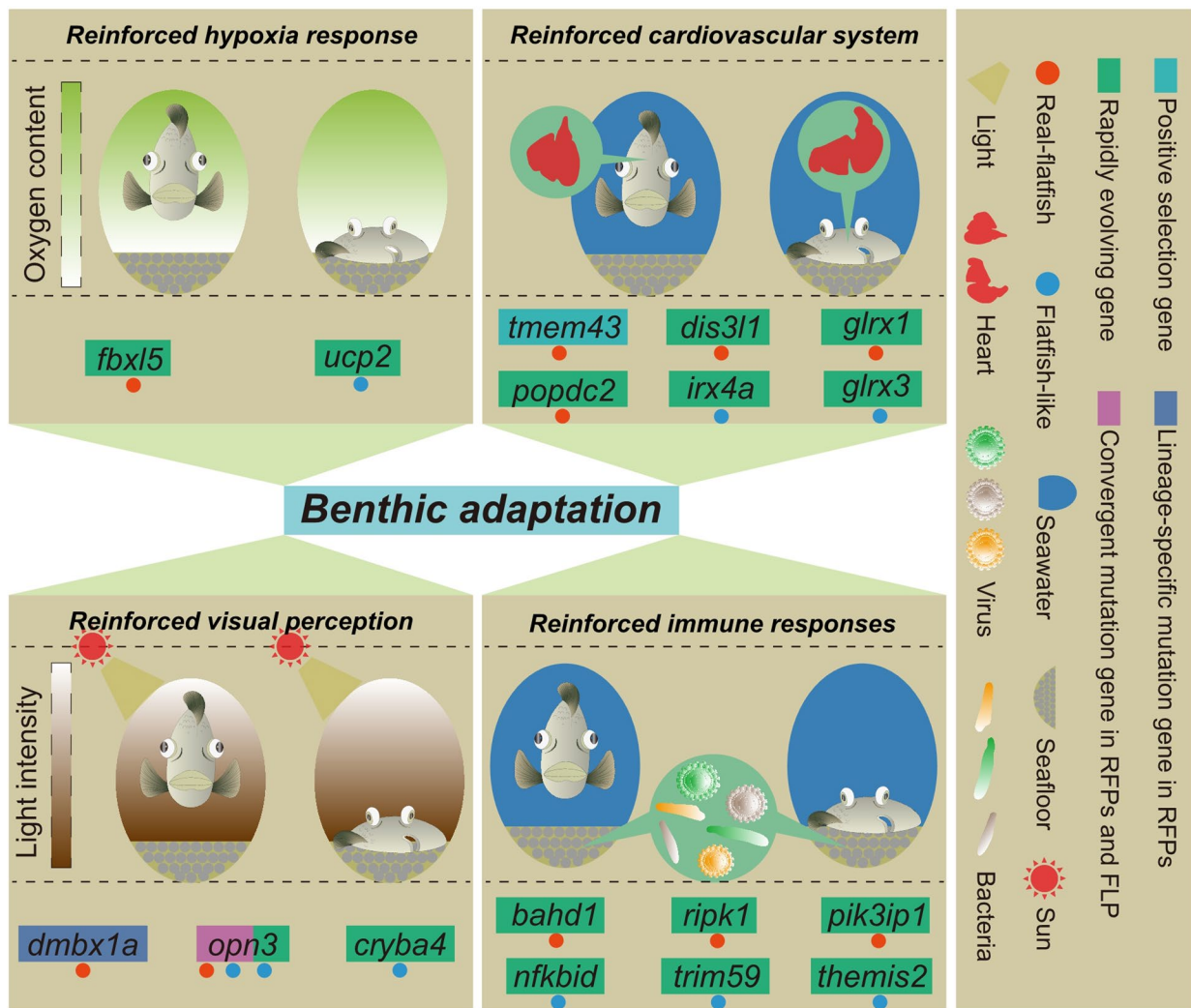
Extended Data Fig. 3 | Genes associated with seafloor colonization of flatfishes. The genes associated with certain seafloor colonization adaptation are shown below in the panel. Genes under positive selection, fast evolution, lineage specific mutation or possessing lineage-specific conserved non-coding elements are marked in different colors. The red and blue dots represent real flatfish Pleuronectoidei and flatfish-like Psettodoidei lineages, respectively. The enlarged red diagrams in the panel of 'reinforced cardiovascular system' represent heart tissues. The enlarged green diagrams in the panel of 'reinforced immune responses' represent bacteria and viruses.

## The relative ratio of left-right axis to total length
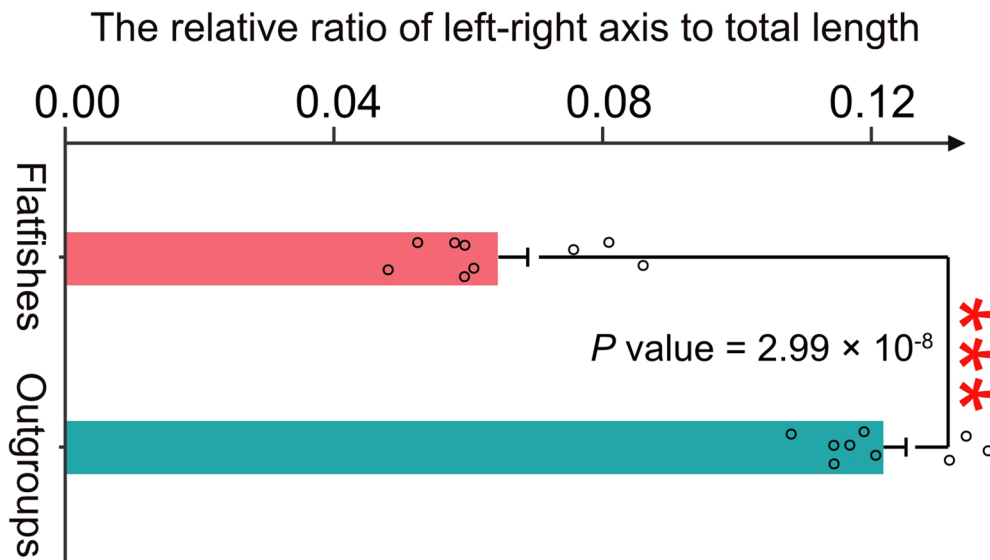
Extended Data Fig. 4 | The relative ratio of left-right axis to total length. The relative ratio of maximum length of left-right axis to the total length was used here to indicate the degree of body pan flatness of fishes. The ratio was measured in three individuals for each species and the data are presented as mean values ± SD. The statistical difference between groups was calculated using Student's *t*-test (two tails) with '***' representing a statistical significance of *P* value < 0.001.
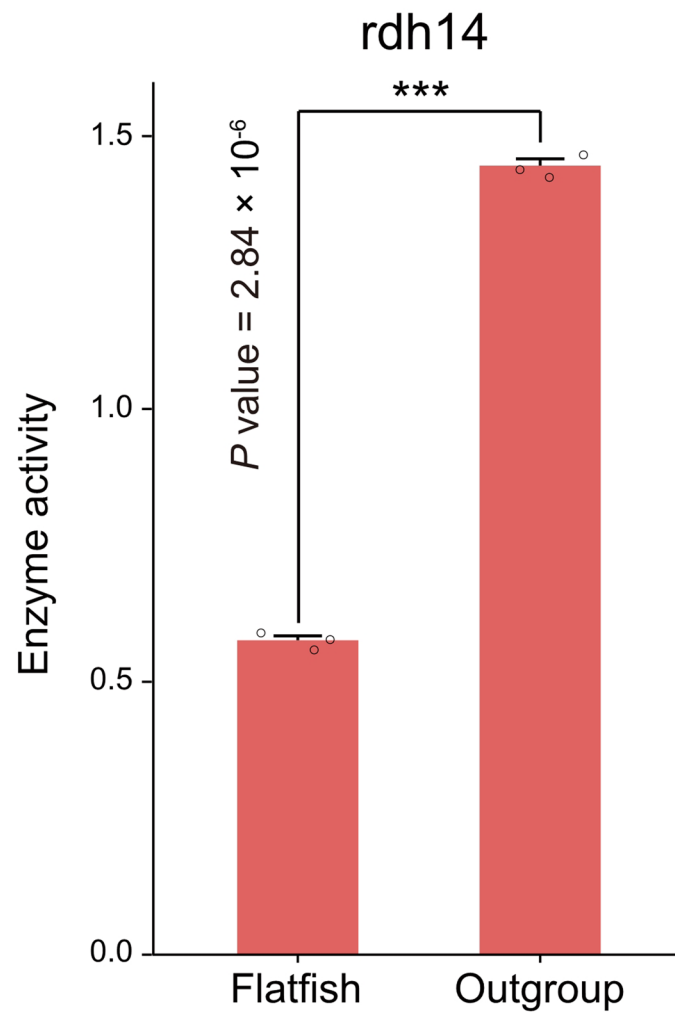
Extended Data Fig. 5 | The catalyzing efficiency of enzyme rdh14 in RFP compared to that of the outgroups. The x-axis represents the RFP specific rdh14 and that of the outgroups. The y-axis represents the measured relative catalyzing efficiency of rdh14. Each experiment was performed with three reaction replicates to determine the mean values ± SD. The distinction of enzyme catalytic activity of RFP compared to non-flatfish teleosts was tested using Student's *t*-test (two tails) with '***' represents a statistical significance of *P* value < 0.001.
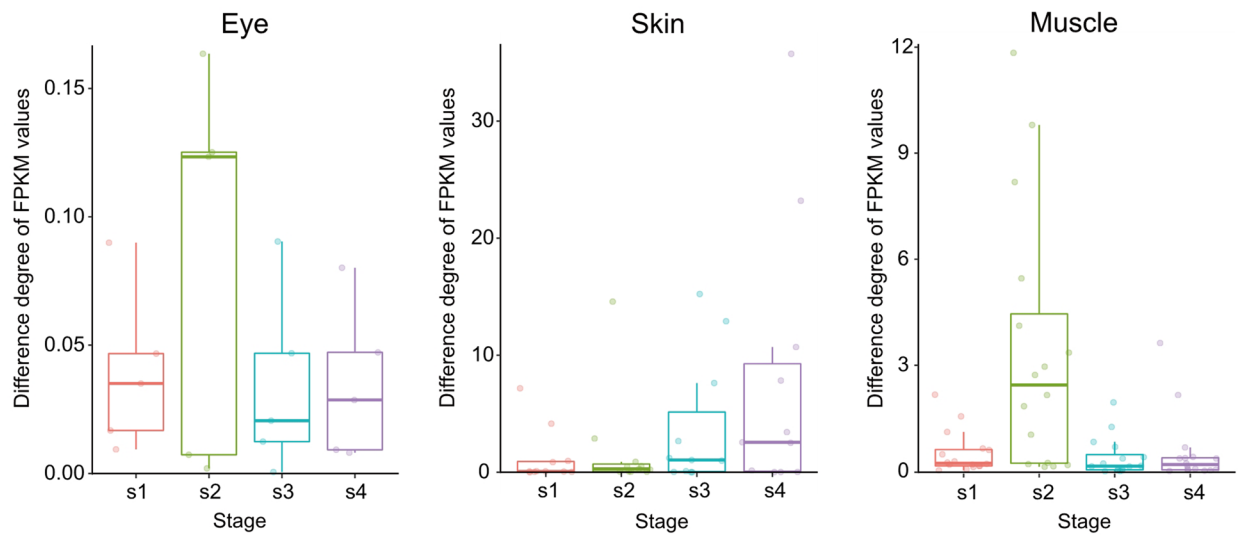
Extended Data Fig. 6 | Asymmetric gene expression in metamorphic flounders. The symbol s1-s4 in the x-axis represents the four developmental stages of flounder, including the pre-metamorphic stage, the pro-metamorphic stage, the metamorphic climax stage, and the post-metamorphic stage, respectively. The y-axis represents the gene expression difference across the left-right axis of flounder, as indicated by FPKM values. The minima, maxima, centre, and the upper and lower bounds of box represent the maximum, minimum, median value, upper and lower quartile, respectively. Each colored dot represents one gene in WNT, RA, or NODAL signal pathway, or gene that associated with pigmentation that are equivalent to those indicated in Fig. 4d in the main text.
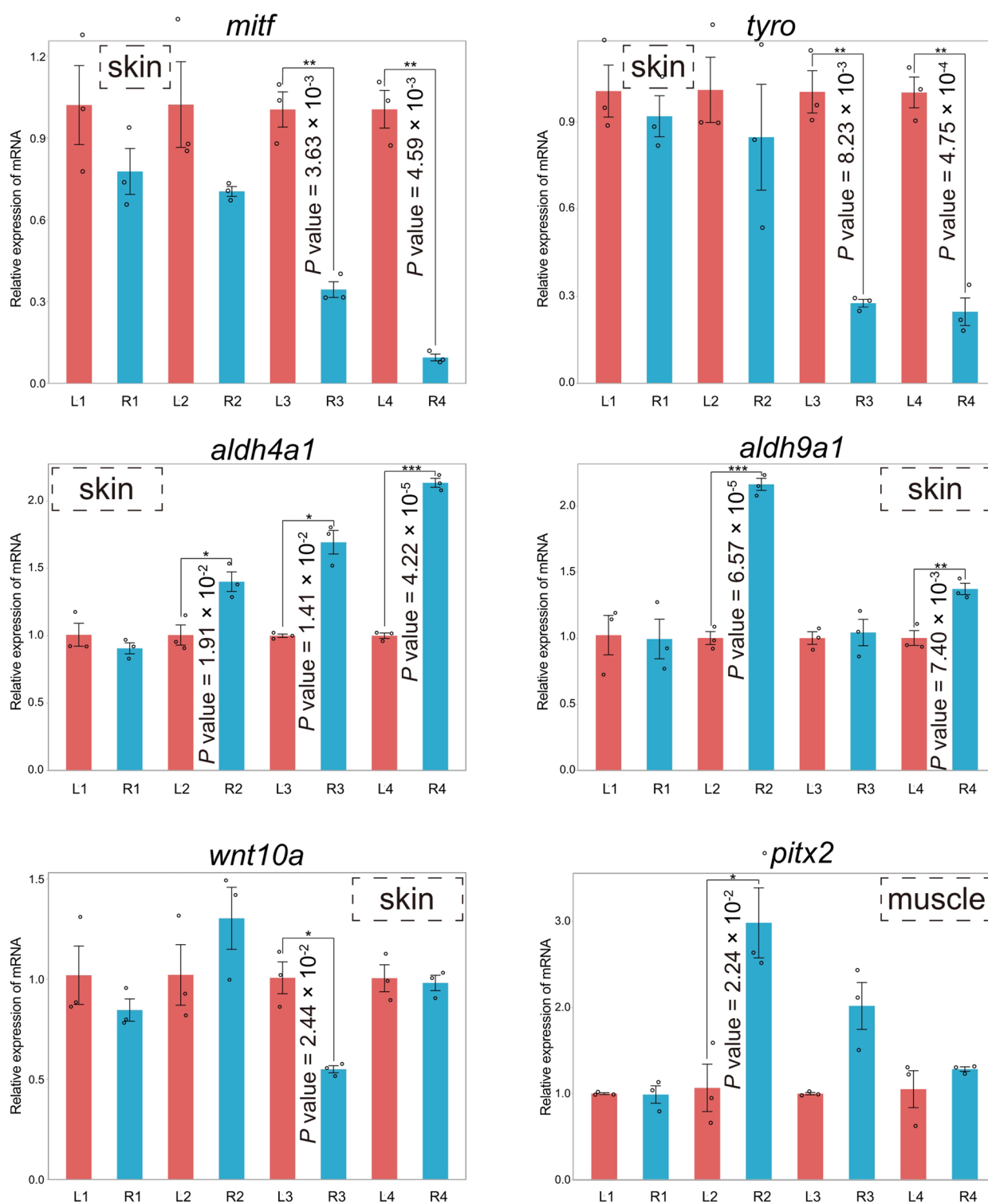
Extended Data Fig. 7 | Asymmetric gene expression confirmed by Real-time quantitative PCR. For each metamorphic time window, three biological replicates were sampled, with each replicate containing tissues from at least 30 individuals, and were used for the RNAs extraction and q-PCR analysis. The numbers 1-4 in the x-axis represent the four developmental stages of flounder, including the pre-metamorphic stage, the pro-metamorphic stage, the metamorphic climax stage, and the post-metamorphic stage. L and R in the x-axis represent tissues on left-side and right-side of the larva, respectively. The y-axis represents the relative expression level of genes compared with internal control of beta-actin and the data are presented as mean values ± SD. The left-right distinction of gene expression profiles in each metamorphic time window was tested using Student's t-test (two tails), and '*', '**', and '***' represents a statistical significance of P value < 0.05, 0.01, and 0.001, respectively.
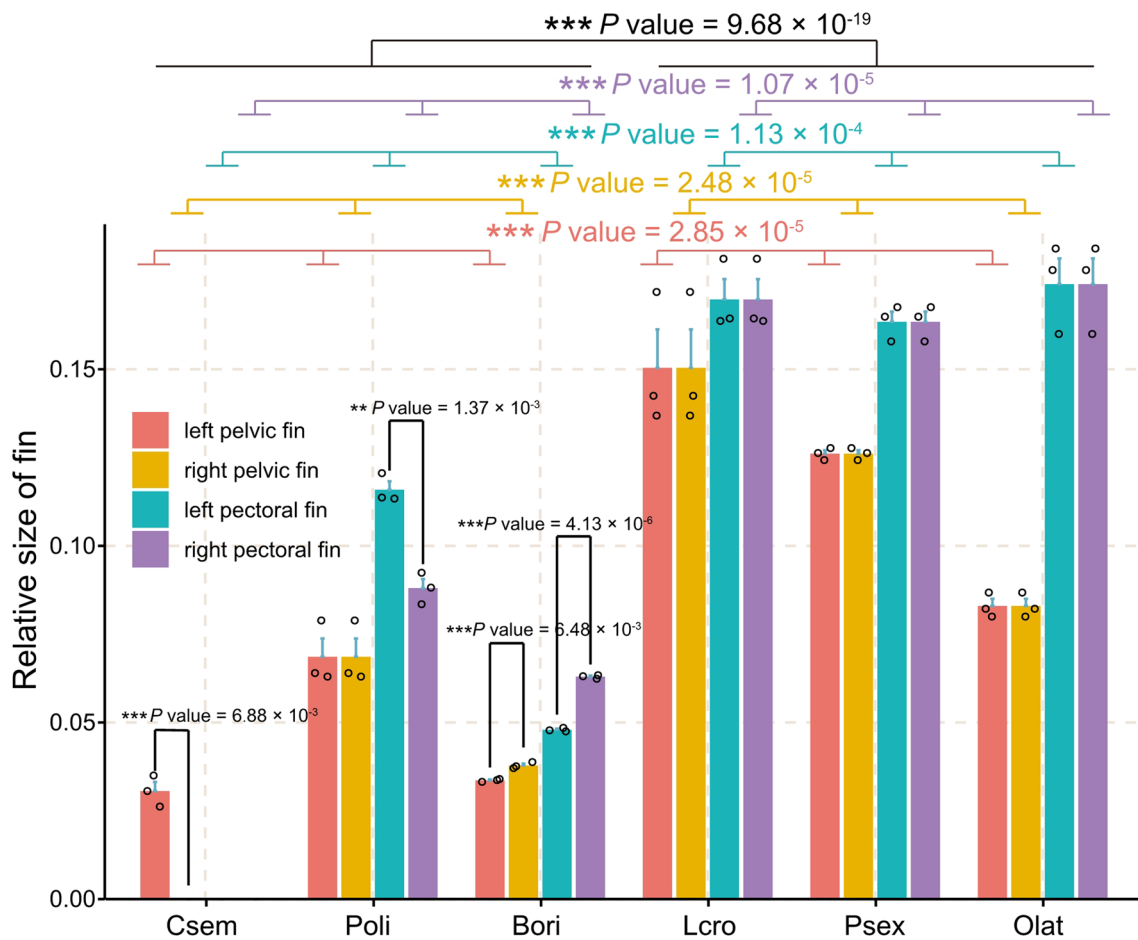
Extended Data Fig. 8 | Relative size of pelvic and pectoral fins in flatfishes compared to non-flatfish species. The x-axis in the panel shows the species with pectoral and pelvic fins measured, and the y-axis shows the relative sizes of fins represented by the ratios of length of pectoral and pelvic fins to the total length of the fish. All the parameters were measured in three individuals for each species and the data are presented as mean values ± SD. The distinction of fin morphology of flatfishes compared to non-flatfish teleosts was tested using Student's *t*-test (two tails) with '**' and '***' representing a statistical significance of *P* value < 0.01 and *P* value < 0.001 respectively. Csem (*Cynoglossus semilaevis*); Poli (*Paralichthys olivaceus*); Bori (*Brachirus orientalis*); Lcro (*Larimichthys crocea*); Psex (*Polydactylus sextarius*); Olat (*Oryzias latipes*).
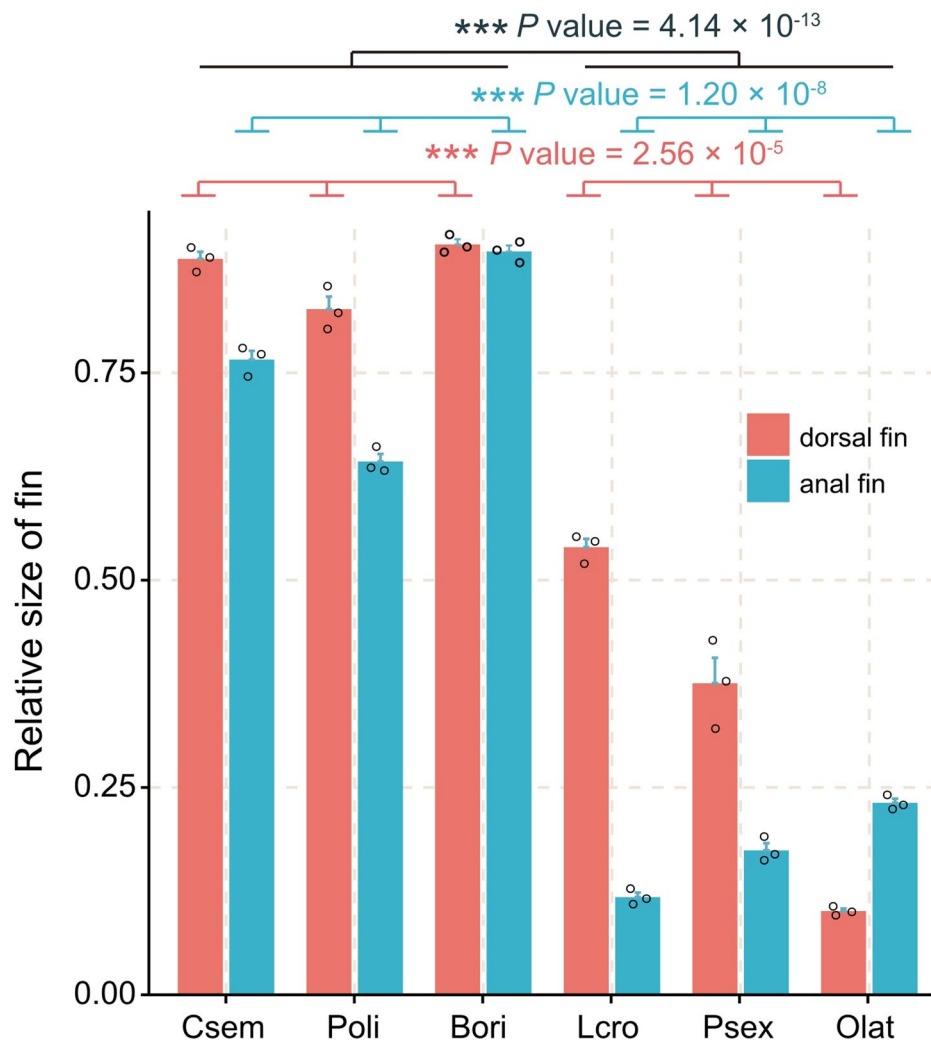
Extended Data Fig. 9 | Relative sizes of dorsal and anal fins in flatfishes compared to non-flatfish species. The x-axis in the diagram shows the species with dorsal and anal fins measured, and the y- axis in the diagram shows the relative sizes of fins represented by the ratios of length of dorsal and anal fins to the total length of the fish. All the parameters were measured in three individuals for each species and the data are presented as mean values ± SD. The distinction of fin morphology of flatfishes compared to non-flatfish teleosts was tested using Student's *t*-test (two tails) with '***' representing a statistical significance of *P* value < 0.001. Csem (*Cynoglossus semilaevis*); Poli (*Paralichthys olivaceus*); Bori (*Brachirus orientalis*); Lcro (*Larimichthys crocea*); Psex (*Polydactylus sextarius*); Olat (*Oryzias latipes*).

Extended Data Fig. 10 | The lineage specific mutated gene of *hoxd12a* in real flatfish species. The sites that showed variation between species are marked in different colors. The fixed variation site between real flatfish Pleuronectoidei species and outgroups are marked with a dashed box.

Corresponding author(s):   Yongxin Li; Xiaoyu Kong; Wen Wang

Last updated by author(s):   Feb 1, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	We list here the versions of all open-source tools used for data analysis.
	1) HiC-Pro (v3.2);
	2) Platanus (v1.2.4);
	3) GapCloser (v1.10);
	4) WTDBG (v1.2.8);
	5) Pilon (v1.21);
	6) Juicer (v1.5);
	7) 3D de novo assembly (v170123);
	8) Bridger (r2014-12-01);
	9) BWA (v0.7.12);
	10) BLAT (v34);
	11) LAST (v802);
	12) Tandem Repeat Finder (v4.04);
	13) RepeatMasker (v4.0.6);
	14) RepeatModeler (v1.0.8);
	15) Augustus (v3.2.1);
	16) GENSCAN (v1.0);
	17) Genewise (v2.2.0);
	18) PASA (v2.1.0);
	19) EvidenceModeler (v1.1.1);
	20) InterProScan (v4.8);

21) BLAST (v2.6.0);
22) OrthoMCL (v2.0.9);
23) MUSCLE (v3.8.31);
24) RAxML (v8.2.9);
25) MPEST (v2.0);
26) OrthoFinder (v2.3.5);
27) PAML (v4.8);
28) LINTRE (Unix version);
29) MEGA (v7.0.26);
30) CAFE (v3.1);
31) Tophat2 (v2.1.1);
32) Cufflinks (v2.2.1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the sequencing data was deposited at the NCBI (Platichthys stellatus: PRJNA592732; Trinectes maculatus: PRJNA592733; Brachirus orientalis: PRJNA592734; Paraplagusia blochii: PRJNA592738; Chascanopsetta lugubris: PRJNA592739; Colistium nudipinnis: PRJNA592742; Pseudorhombus dupliocellatus: PRJNA592743; Polydactylus sextarius: PRJNA592744; Toxotes chatareus: PRJNA592745; Psettodes erumei: PRJNA592748; Paralichthys olivaceus: PRJNA632737). Besides, the source data of the fish photos were deposited at the Figshare database (https://doi.org/10.6084/m9.figshare.13664201.v1).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	DNA were extracted from one sample for each species and sequenced at least 60X for the subsequent genome assembly. One sample for each species was used for genome sequencing, which is the routines for genome analyses in many taxon (Chen et al, 2014, doi: 10.1038/ng.2890; Shao et al, 2017, doi: 10.1038/ng.3732). For transcriptome sequencing and q-PCR analysis in Paralichthys olivaceus, we actually sampled muscle, eye and skin from both sides of at least 30 individual larvae and then respectively pooled each type of tissues together to conduct RNA extraction. We repeated three times from different batch of larvae for the sample collection as three biological replicates. Body and fin morphology, fat content of each species were measured in three individuals. For catalytic activity assay of enzyme, we conducted three reaction replicates to determine the mean +/- SD and calculate the statistical significance of differences between groups with a p value.
Data exclusions	No data were excluded.
Replication	For transcriptome sequencing and real-time PCR analysis, the analyses were conducted on three replicates for each tissue/stage (Paralichthys olivaceus). For body and fin morphology, fat content, the analyses were conducted on three individuals for each species. For catalytic activity assay of enzyme, we conducted three reaction replicates. All these experiments were technically repeated to ensure the reliability of the results.
Randomization	The samples for genomic DNA sequencing of each species were all randomly picked. The flounder larvae per development stage were also randomly picked to test the expression profiles of genes possibly involved in the asymmetrical development of body plans. The samples were also randomly selected to analyze the fat content, body and fin morphology of flatfishes, compared to outgroup teleosts.
Blinding	Investigators were blinded to group allocations during the specimen sampling, DNA/RNA extraction, library construction and sequencing process, and other phenotypical and biochemical analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☐ ☒	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Animals and other organisms

Policy information about <u>studies involving animals</u>; <u>ARRIVE guidelines</u> recommended for reporting animal research

Laboratory animals	The flounder larvae used for the transcriptome analyses were collected from a flounder hatchery station located in Qingdao city, Shandong province, China.
Wild animals	Trinectes maculatus, Toxotes chatareus, and Colistium nudipinnis were collected from America, Thailand, and Australia, respectively, through laboratory to laboratory sample exchange programme. For other fish species, all the samples were collected by trawling from China during fishery resource surveys launched by South China Sea Institute of Oceanology, Chinese Academy of Sciences, and Zhejiang Ocean University. All the samples were collected dead from these survey activity and transported at low temperature from sampling sites to the laboratory and stored in -80 °C. For transcriptome analysis, adult fish and metamorphic flounder larvae were anaesthetized using MS-222 immersion before tissues were sampled. The adult fish and metamorphic larvae of Paralichthys olivaceus were used in the experiments and this specific study did not involve the sex of the species and so not identified, which has been clearly described in main text and supplementary information.
Field-collected samples	For genome sequencing, fat content, and phenotypical analysis, samples were collected dead from fishery resource surveys activities or through laboratory to laboratory sample exchange programme. For transcriptome analysis, adult fish and metamorphic flounder larvae were transported alive in aired seawater in low temperature and anaesthetized immediately after arrival using MS-222 immersion before tissues were sampled.
Ethics oversight	This study was approved by Institutional Animals Care and Use Committee of Zhejiang Ocean University and by Experimental Animal Management and Ethics Committee of South China Sea Institute of Oceanography, Chinese Academy of Sciences.

Note that full information on the approval of the study protocol must also be provided in the manuscript.