Hu et al. *Zool. Res.* 2025, 46(3): 661−674
https://doi.org/10.24272/j.issn.2095-8137.2024.388
https://cstr.cn/32248.14.zr.2024.388

ZR | Zoological Research

Article Open Access

Evolutionary balance between genomic conservation and coral reef adaptation in the yellow boxfish (*Ostracion cubicus*)

Shui-Mu Hu[1,2,#], Zhi-Xiong Zhou[1,#], Jun-Yi Yang[1], Zhou Jiang[1], Fei Pu[1], Qing-Ming Qu[3], Tao Zhou[2,*], Peng Xu[1,2,*]

[1] *State Key Laboratory of Mariculture Breeding*, *College of Ocean and Earth Sciences*, *Xiamen University*, *Xiamen*, *Fujian* 361102, *China*
[2] *Shenzhen Research Institute of Xiamen University*, *Xiamen University*, *Shenzhen*, *Guangdong* 518000, *China*
[3] *School of Life Sciences*, *Xiamen University*, *Xiamen*, *Fujian* 361102, *China*

ABSTRACT

The yellow boxfish (*Ostracion cubicus*) exhibits a combination of derived morphological traits specialized for coral reef environments and ancestral characteristics, including a fused dermal plate. Contradictory evolutionary evidence hinders true classification of *O. cubicus*. To clarify its evolutionary position within Tetraodontiformes, a chromosome-level genome assembly was generated, representing the most contiguous and complete genome to date for this lineage. Notably, *O. cubicus* possessed the largest genome within the order Tetraodontiformes, primarily due to extensive transposable element expansion. Phylogenetic analysis based on 19 whole genomes and 131 mitochondrial genomes resolved Tetraodontiformes into three major sister groups (Ostraciidae-Molidae, Tetraodontidae, and Balistidae-Monacanthidae). Comparative genomic evidence indicated that *O. cubicus* diverged early from the common ancestor of modern Tetraodontiformes and retained the highest number of *HOX* genes among surveyed taxa. Although overall genomic architecture was largely conserved, certain genetic and environmental changes may have contributed to its phenotypic adaptations, including climate cooling during the Miocene-Pliocene Transition, recent DNA and long interspersed nuclear element (LINE) transposon bursts, lineage-specific chromosomal rearrangements, and gene family expansion. Many positively selected genes and rapidly evolving genes were associated with skeletal development, including *bmp7*, *egf7*, and *bmpr2*. Transcriptomic comparisons between carapace and tail skin revealed various candidate genes and pathways related to carapace formation, such as *postn*, *scpp1*, and components of the TGF-β signaling pathway. A derived amino acid substitution in *eda*, coupled with protein structural modeling, suggested potential molecular convergence in dermal plate formation among teleosts. These findings provide novel insights into the genomic and developmental basis of carapace evolution and coral reef-adaptation in *O. cubicus*, offering a strong case for evolutionary balance between genomic conservation with regulatory innovation to achieve coral reef specialization.

Keywords: *Ostracion cubicus*; Comparative genomics; Evolutionary genomics; Dermal carapace; Convergent evolution

INTRODUCTION

The order Tetraodontiformes exhibits remarkable phenotypic diversification driven by adaptive radiation, with distinct defense strategies evolving across its major lineages. These include dermal armor (Ostracioidei), hypertrophied spines (Diodontidae), potent chemical defenses (Ostracioidei and Tetraodontoidei), body inflation (Tetraodontoidei), and extreme body enlargement (Moloidei) (Tyler & Santini, 2002). Ostraciidae, a lineage encompassing Ostraciidae and Aracanidae, is one of the most morphologically specialized clades. Commonly referred to as boxfish or trunkfish, members of Ostraciidae are characterized by a rigid, full-length bony carapace, square or triangular body shape, and the production of pahutoxin, a potent ichthyotoxin (Hove et al., 2001). The family includes approximately 30 extant species inhabiting tropical coral reef systems (Holcroft, 2005). *Ostracion cubicus*, the yellow boxfish, is a representative species of the Ostraciidae family, known for its unique cubic body form and vivid coloration. Paleontological data suggest deep evolutionary origins of these traits. The extinct *Plectocretacicus clarae*, the oldest known Tetraodontiformes species from the Tethys region, has been shown to share a similar body shape and armor with boxfish (Arcila et al., 2015). Dermal armor represents a conserved defense strategy

Received: 03 January 2025; Accepted: 10 February 2025; Online: 11 February 2025

Foundation items: This study was supported by the Basic and Applied Basic Research Foundation of Guangdong Province (2020A1515110882) and National Science Fund for Distinguished Young Scholars (32225049)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: zt@xmu.edu.cn; xupeng77@xmu.edu.cn

observed across both fossil and extant lineages, including jawless fish, sea horses, sea robins, and seamoths (Janvier, 2008; Yang et al., 2015). While rigid armor may constrain swimming speed and reduce agility, its protective benefits favor survival in structurally complex habitats such as benthic sediment and coral reefs, where concealment is important. In addition to their robust carapace, boxfish also display a complex array of specialized coral reef adaptations, including box-like body shape, exceptional maneuverability, juvenile aposematism, and rapid toxin synthesis (Gordon et al., 2000; Kalmanzon & Zlotkin, 2000; Marshall et al., 2019; Van Wassenbergh et al., 2015). These novel phenotypes contribute to their ecological success and limit natural predation within reef ecosystems. However, the genomic mechanisms underlying the evolution of these phenotypic adaptations in *O. cubicus* remain unclear.

Phylogenetic relationships within Tetraodontiformes, the order to which *O. cubicus* belongs, remain contentious (Holcroft, 2005; Matsuura, 2015; Santini et al., 2013; Yamanoue et al., 2008). This controversy largely stems from two factors: (i) extensive morphological diversity and character convergence across Tetraodontiformes obscure classification boundaries based on traditional morphological methods, and (ii) different molecular markers and analytical methods frequently yield conflicting topologies, complicating efforts to resolve deep evolutionary relationships. Precise inferences about genome evolution are highly dependent on high-quality genomic resources and a robust phylogenetic framework. The recent release of chromosome-level genomes for important Tetraodontiformes species, such as *Takifugu rubripes* (Aparicio et al., 2002), *Takifugu flavidus* (Gao et al., 2014), *Takifugu bimaculatus* (Zhou et al., 2019), *Takifugu obscurus* (Kang et al., 2020), *Mola mola* (Pan et al., 2016), *Tetraodon palembangensis* (Zhang et al., 2021), and *Thamnaconus modestus* (Yang et al., 2019), enables genome-wide comparisons that can resolve incongruences in phylogenetic signal and refine the evolutionary placement of *O. cubicus*. Comparing whole genomes across different species also permits systematic investigation into whether conserved or non-conserved genetic or environmental factors are associated with the adaptive diversification of this morphologically unique species.

The evolution of dermal armor in fish lineages provides a classic example of convergent evolution (Yang et al., 2013). Bony plates, a specialized form of mineralized scale, have evolved independently across multiple bony fish clades, including Acipenseriformes, Lophiiformes, Syngnathiformes, Tetraodontiformes, Trachiniformes, Diodontiformes, and Ostraciiformes (Lemopoulos & Montoya-Burgos, 2021). Among these, yellow boxfish are characterized by a robust carapace covering their entire body, making them excellent subjects for studying the development and functional role of dermal armor in reef-dwelling fish species. This specialized architecture confers distinct adaptive advantages in structurally complex reef environments, potentially enhancing defense against predators. Evolutionary reduction of primitive spiny scales in ancestral lineages resulted in the retention and progressive modification of basal scale elements, which gradually evolved into the bony plates observed in modern boxfish (Yang et al., 2015). The armor consists of many polygonal scutes, each composed of a highly mineralized surface plate and basal collagen base (Garner et al., 2020). At scute boundaries, mineralized plates exhibit a tooth-like suture morphology, while the underlying collagen layers are interconnected by bridging collagen fibers. Biomechanical testing has shown that the armor of boxfish can resist penetrative and compressive forces, underscoring its defensive efficacy within predator-rich ecosystems (Yang et al., 2015). Although prior studies have characterized the mechanical strength, structural stability, and chemical composition of dermal plates (Garner et al., 2020; Porter et al., 2017; Qiu et al., 2023), the genetic basis of their formation remains largely unresolved.

This study investigated the genomic foundations of phenotypic adaptability in *O. cubicus*, focusing on its distinctive morphological and ecological traits. Using a newly assembled chromosome-level genome, alongside comparative genomic and transcriptomic analyses, the research aimed to resolve key evolutionary questions: (i) How are Tetraodontiformes species phylogenetically related based on whole-genome comparisons? (ii) Which genomic features and specific loci are associated with phenotypic adaptations in *O. cubicus*? (iii) Does the evolution of the *O. cubicus* carapace provide insight into the molecular basis of bony plate convergence across teleost lineages?

MATERIALS AND METHODS

Sourcing and sequencing

The juvenile *O. cubicus* specimen used in this study was obtained from Hainan Province and reared in an artificial coral reef system at the College of Ocean and Earth Sciences, Xiamen University. Tissues including muscle, liver, brain, kidney, intestine, spleen, heart, skin, and eye were collected and immediately frozen in liquid nitrogen, then stored at −80℃. High-quality genomic DNA was extracted from muscle tissue using a QIAGEN DNeasy Blood & Tissue Kit (Qiagen, China) and used for library construction. Illumina sequencing was performed on the HiSeq X Ten platform (2×150 bp, 350 bp insert size) using a TruSeq Nano DNA Library Prep Kit (Illumina, USA), generating 66.21 Gb of data. PacBio sequencing was conducted with high-molecular-weight gDNA (>20 kb fragments) on a PromethION sequencer (Oxford Nanopore, UK), yielding 69.39 Gb of data. Hi-C libraries were prepared using formaldehyde fixation and DpnII digestion (New England Biolabs, USA), generating 87.94 Gb of raw reads on the Illumina HiSeq X Ten platform (Illumina, USA). For RNA sequencing, tail skin and carapace samples were dissected and collected from three *O. cubicus* individuals. Total RNA was extracted using TRIzol reagent (Invitrogen, USA) and sequenced on the Illumina HiSeq X Ten platform (Illumina, USA), generating 32.61 Gb of clean reads for gene prediction and transcriptomic analysis. Data sources are listed in Supplementary Table S1. All animal treatments and procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of Xiamen University (Approval No. XMULAC20240215).

Genome assembly, quality assessment, and annotation

Reads from the Illumina, PacBio, and Hi-C libraries were integrated for genome assembly and annotation. Illumina reads were used for initial genome survey analysis, PacBio reads enabled contig assembly, and Hi-C reads facilitated chromosome-level scaffolding. Genome size and heterozygosity were estimated from 17-mer frequency distributions using Jellyfish (v.2.3.0) and established methods (Marçais & Kingsford, 2011) (Supplementary Figure S1). Contigs were assembled using Canu (v.2.0) (Koren et al.,

2017), then polished sequentially with Racon (v.1.2.1) (Vaser et al., 2017) and Pilon (v.1.22) (Walker et al., 2014). Chromosome anchoring was performed using Hi-C reads with Juicer (v.1.6) (Durand et al., 2016) and 3D-DNA (v.180419) (Dudchenko et al., 2017), followed by manual refinement using Juicebox Assembly Tools (https://github.com/theaidenlab/juicebox). The final assembly comprised 25 chromosomes. Genome accuracy and completeness were evaluated using Compleasm (v.0.2.6) (Huang & Li, 2023) and BUSCO (v.5.6.1) (Manni et al., 2021) with the *Actinopterygii*_odb10 lineage database to assess the presence of single-copy orthologs.

Repeat elements were annotated using both *de novo* and homology-based methods. RepeatModeler (v.2.0.1) (Flynn et al., 2020) and LTR_Finder (v.1.07) (Xu & Wang, 2007) were employed to generate a *de novo* repeat library, which was combined with Repbase and used by RepeatMasker (v.4.1.0) (Tarailo-Graovac & Chen, 2009) for classification. Unclassified elements were annotated with TEclass (v.2.1.3) (Abrusán et al., 2009). Built-in RepeatMasker (v.4.1.0) scripts, buildSummary.pl, calcDivergenceFromalign.pl, and createRepeatLandscape.pl, were used to summarize transposable elements (TEs), determine Kimura divergence values of TEs, and illustrate TE landscapes, respectively. The Kimura two-parameter method estimated TE insertion age by comparing nucleotide distances between TE copies. Tandem repeats were identified using Tandem Repeats Finder (v.4.0.9) (Benson, 1999), and all repetitive regions, except for tandem repeats, were soft-masked for protein-coding gene annotation. Non-coding RNA (ncRNA) annotation was conducted using tRNAScan-SE (v.1.3.1) (Lowe & Eddy, 1997) for tRNA prediction and RNAmmer (v.1.2) (Lagesen et al., 2007) for rRNA prediction. Other ncRNAs were predicted by searching the Rfam database (http://eggnogdb.embl.de/).

Protein-coding genes were predicted through *ab initio*, homology-based, and transcriptome-assisted methods. RNA-seq reads were assembled using Trinity (v.2.10.0) (Rahner, 2001), and homologous coding sequences from related species were obtained from the European Nucleotide Archive and aligned using GenomeThreader (v.1.7.0) (Gremme et al., 2005). *Ab initio* gene prediction was performed based on transcripts assembled from RNA-seq and known genes from related species using Braker2 (Brůna et al., 2021). RNA-seq data were aligned to the *O. cubicus* genome using HISAT2 (v.2.2.1) (Kim et al., 2015) and assembled with StringTie (v.2.2.1) (Pertea et al., 2016), followed by open reading frame (ORF) prediction with TransDecoder (v.5.5.0, https://github.com/TransDecoder/TransDecoder). EvidenceModeler (v.2.1.0) was used to create a comprehensive gene set, which was further annotated for protein-coding gene structure using PASA (v.2.4.1) (Haas et al., 2008). Functional annotation was performed by aligning predicted genes against Swiss-Prot, InterPro, and NR databases using Diamond (v.2.0.6) (Buchfink et al., 2021), with an *E*-value<1e–5. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations were performed using InterProScan (v.5.68) (Jones et al., 2014) and the KEGG Automatic Annotation Server (KAAS, https://www.genome.jp/tools/kaas/).

Phylogenetic relationships and population history construction

Single-copy orthologous genes shared between *O. cubicus* and 18 other representative species were identified using OrthoFinder (v.2.5.4) (Emms & Kelly, 2019) based on gene families constructed from the protein sequences of all species, then aligned using MUSCLE (v.3.8.31) (Edgar, 2010). A combined ultralong sequence was created from all translated coding DNA alignments for maximum-likelihood (ML) phylogenetic tree construction using RAxML-NG (Kozlov et al., 2019) and the PROTGAMMAAUTO model with 500 rapid bootstrap replicates. Divergence time was estimated using MCMCTREE in the PAML package, calibrated against fossil-based molecular clock constraints retrieved from the TimeTree database (Álvarez-Carretero et al., 2022), using fourfold degenerate third codon transversions (4DTV) for rate modeling. To complement genome-based inference, mitochondrial genomes from 131 species were aligned using MUSCLE (v.3.8.31) (Edgar, 2010) and analyzed using RAxML-NG (Kozlov et al., 2019) with the GTRGAMMA model and 500 rapid bootstrap replicates. Demographic histories of *O. cubicus*, *T. rubripes*, *M. mola*, and *Thamnaconus modestus* were inferred using the Pairwise Sequentially Markovian Coalescent (PSMC) method (https://github.com/lh3/psmc) with the parameters: N=25, t=15, r=5, p=4+25*2+4+6. The neutral mutation rate (μ) was estimated using r8s. Effective population size (Ne) dynamics were normalized by dividing Ne at each time point by the maximum value for the same species across the full time scale.

Comparative and evolutionary genomic analyses

To reconstruct ancestral chromosomal architecture within Tetraodontiformes, the *O. cubicus* genome was used as the reference. *Lophius litulon* served as the main outgroup, while *Gasterosteus aculeatus* and *Oryzias latipes* served as secondary outgroups. Pairwise alignments were generated using LASTZ (v.1.04.00) (Harris, 2007) with the parameters: T=2, C=2, H=2 000, Y=3 400, L=6 000, K=2 200. The alignment results were then converted into UCSC “chain” and “net” formats using UCSC pipeline tools (axtChain, chainSort, chainPreNet, chainNet, and netSyntenic) with default parameters. DESCHRAMBLER (Kim et al., 2017) was employed to construct ancestral chromosomes at a resolution of 300 kb. MCScanX (Wang et al., 2012) and JCVI (v.1.4.18) (Tang et al., 2024) were used for detailed collinearity analysis between *O. cubicus* and other species (*Oryzias latipes*, *T. rubripes*, and *Thamnaconus modestus*), and visualized through collinearity plots.

Gene families were inferred using OrthoFinder (v.2.5.4) after filtering protein sequences shorter than 30 amino acids (Emms & Kelly, 2019). Expansions and contractions of gene families in *O. cubicus* were identified using CAFÉ (v.4.2.1) (Han et al., 2013). Members of the *SCPP* and *HOX* gene families were extracted from orthogroups for phylogenetic reconstruction using ClustalW (Larkin et al., 2007) and visualized with ML trees and gene structure maps using the R package gggenome. Positively selected genes (PSGs) and rapidly evolving genes (REGs) were identified using PRANK-MSA (v.140110) (Löytynoja, 2014) with the parameters gaprate=0.025 and gapext=0.75 for coding sequence alignment of each homologous group. To estimate selective constraints, the dN/dS ratio (ω) was calculated using PAML (v.4.4b) (Yang, 2007). REGs were identified using a branch model that allows different rates for different branches, with the null model assigning the same ratio to all branches. PSGs were identified using the branch-site model, which allows sites to be under positive selection on the foreground branch, while the null mode restricts sites to neutral evolution or purifying

selection.

Comparative transcriptomic analysis

Clean RNA-seq reads from *O. cubicus* skin and carapace were aligned to the *O. cubicus* reference genome using HISAT2 (v.2.2.1) (Kim et al., 2015). Transcript assembly and quantification were performed with StringTie (v.2.2.1) (Pertea et al., 2016) and gene expression levels were calculated in fragments per kilobase of exon model per million mapped reads (FPKM). Differentially expressed genes (DEGs) were identified based on the following criteria: *Q*-value<0.05, *P*-value<0.05, $|\log_2(\text{fold change})|>2$. Functional annotation of DEGs was performed using GO and KEGG pathway analyses with OmicShare (www.omicshare.com/tools), with enrichment significance set at *P*<0.05. DEGs were visualized using heatmaps and volcano plots generated in R.

Alternative splicing (AS) events and corresponding genes were identified using rMATS (v.4.2.0) (Shen et al., 2014), which categorized AS events into five types: alternative 3' splice site (A3SS), alternative 5' splice site (A5SS), skipped exon (SE), mutually exclusive exon (MXE), and retained intron (RI). Differential AS (DAS) events between carapace and tail skin were defined by a false discovery rate (FDR) threshold of <0.05. Protein-protein interaction (PPI) networks for DEGs and DAS genes were constructed using STRING (Szklarczyk et al., 2021), with *Danio rerio* as the reference species, and Cytoscape (v.3.10.0) (Shannon et al., 2003). The top 5% of genes in the interaction network were considered hub genes using the MCC algorithm in the cytohubba plugin of Cytoscape.

Convergent evolutionary analysis of *eda*

Orthologous Eda protein sequences from 13 species —six bearing dermal plates (*O. cubicus*, *Odonus niger*, *G. aculeatus*, *Hippocampus erectus*, *Hippocampus comes*, and *Atractosteus spatula*) and seven without bony plates (*G. aculeatus*, *Thamnaconus modestus*, *Tetraodon nigroviridis*, *M. mola*, *Ictalurus punctatus*, *D. rerio*, and *T. rubripes*) —were aligned using MAFFT (v.7.453) (Rozewicki et al., 2019). Convergent amino acid residues were identified and visualized using Jalview (v.2.11.3.3) (Waterhouse et al., 2009).

The genetically mobile domains of 13 protein sequences were predicted and visualized using SMART (http://smart.embl-heidelberg.de/). Tertiary structures of each Eda protein were predicted by AlphaFold-Colab with default parameters and no templates (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb). PyMOL (v.3.0) (Delano, 2002) was used to visualize all protein structures and determine root mean square deviation (RMSD) values between predicted protein structures by pairwise comparison, generating a 13×13 matrix for clustering and visualization using the R package pheatmap. Protein stability was assessed using FoldX (v.5.0) (Delgado et al., 2019); the ΔG of folding was computed using the Stability module, and mutational impacts at convergent residues were evaluated via the BuildModel module.

Structural imaging of *O. cubicus*

Detailed anatomical visualization of the *O. cubicus* carapace was carried out using micro-computed tomography (μ-CT) with a SkyScan 1076 μ-CT scanner (Bruker, Belgium) under the following conditions: aluminum filter: 0.5 mm; isotropic voxel size: 9 μm; electric potential: 70 peak kV (kVp); current: 200 μA.

RESULTS

Chromosome-level genome assembly and annotation of *O. cubicus*

A high-quality reference genome of juvenile *O. cubicus* was constructed using a combination of single-molecule real-time sequencing (PacBio Sequel Ⅰ), paired-end sequencing (Illumina HiSeq X Ten), and Hi-C sequencing (Supplementary Figure S1 and Table S2). Initial contigs were assembled from PacBio reads and subsequently polished with Illumina data, yielding an 864.71 Mb genome with contig N50 of 4.45 Mb (Supplementary Table S3). Hi-C data were then used to anchor the contigs into chromosomes, yielding a final assembly of 837.34 Mb, with 96.77% of bases successfully anchored to 25 chromosomes (2n=50) (Supplementary Figure S1 and Table S4). Assembly quality was validated using BUSCO, which identified 98.5% complete homologs, demonstrating high accuracy of the genome assembly (Supplementary Table S5). In addition, a 16.49 kb mitochondrial genome encoding 13 protein-coding genes was assembled (Figure 1B). Repeat annotation revealed that TEs accounted for 43.77% of the genome, with DNA transposons comprising the largest proportion (23.1%, 199.74 Mb) (Supplementary Table S6). Gene prediction, combining homology-based, *de novo*, and RNA-seq approaches, identified 24 987 protein-coding genes, comparable to that in other related species (Supplementary Figure S2 and Table S7). Of these, 94.97% (23 731 genes) were either supported by homologs in public databases (SwissProt, Nr, Pfam, KEGG, and GO) or assembled RNA-seq transcripts (Supplementary Table S8).

Genome expansion in *O. cubicus* driven by TE proliferation

The *O. cubicus* genome was markedly larger than those of all other Tetraodontiformes species analyzed, ranging from 214% to 258% greater in size, primarily due to its higher TE and intron contents (Figure 1C, D). As major components of vertebrate genomes, TEs play a fundamental role in shaping genome architecture through episodic amplification (Chapman et al., 2010; Talla et al., 2017). Comparative genomic analysis revealed extensive expansion of both Class Ⅱ DNA transposons and Class Ⅰ retrotransposons, including long interspersed nuclear elements (LINEs), long terminal repeat (LTRs), and short interspersed nuclear element (SINEs), in *O. cubicus*, with DNA transposons nearly twice as long as those in related Tetraodontiformes genomes (Figure 1E). Kimura distance-based divergence profiling identified a recent TE burst event in *O. cubicus*, primarily between substitution levels 2 to 5 (Supplementary Figure S3), dominated by specific DNA transposon and LINE families. These elements, including DNA/TcMar-Tc1, DNA/PIF-Harbinger, LINE/L1, and LINE/CR1, were evenly distributed across 25 chromosomes (Supplementary Figure S4). Moreover, a conserved DNA transposon amplification event was detected in both Tetraodontiformes species and *Antennarius maculatus*, with peak activity corresponding to Kimura substitution levels between 15 and 20 (Supplementary Figure S5).

Phylogeny and population dynamics of Tetraodontiformes

To resolve phylogenetic relationships among major Tetraodontiformes species, phylogenies were constructed using 2 098 one-to-one orthologous genes from 19 species

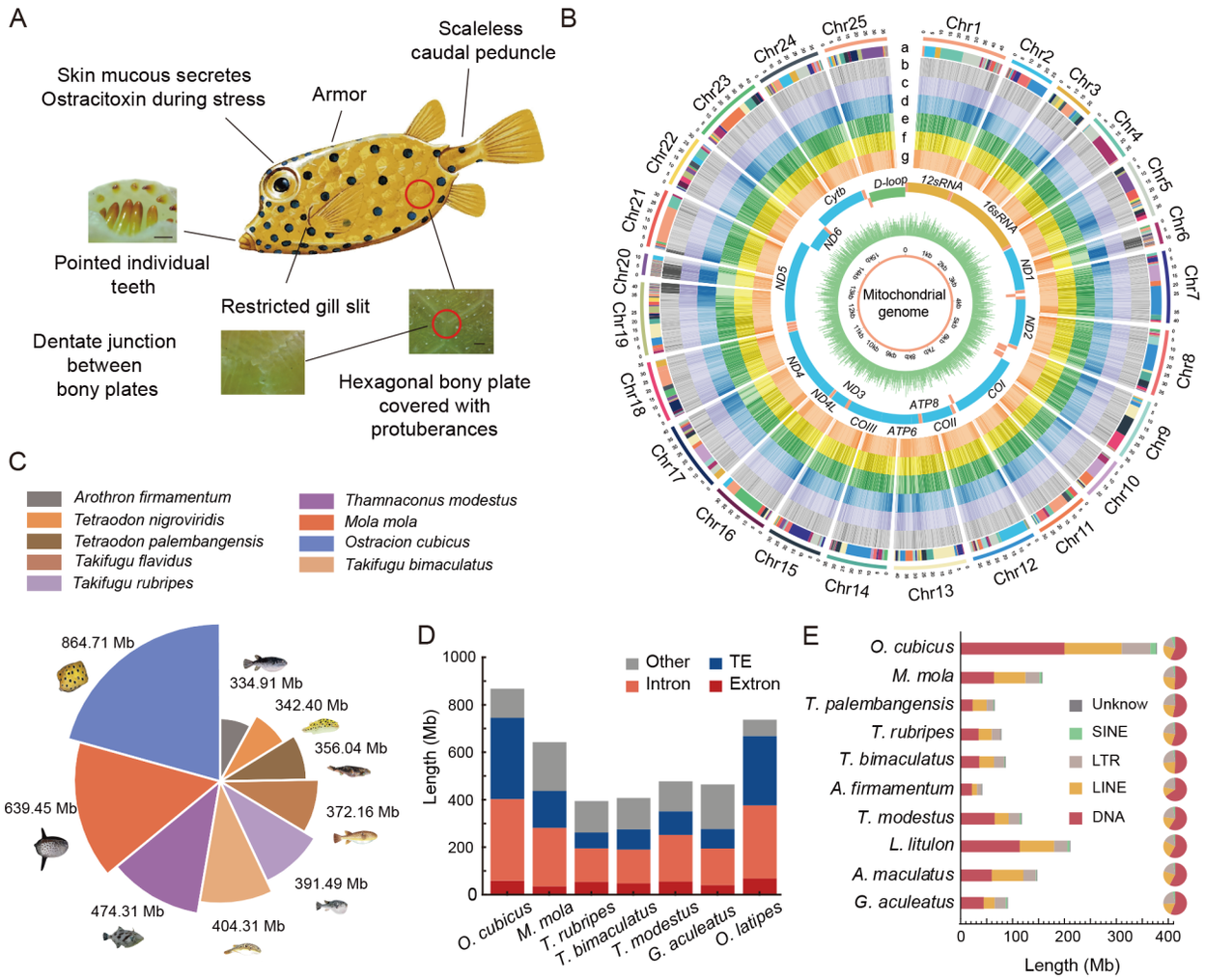

Figure 1 Unique phenotypic and genomic characteristics of *O. cubicus*

A: Schematic of unique yellow boxfish phenotypes. B: Circos plot of yellow boxfish genome. From outside to inside, circle (a) represents scaffold arrangement on each chromosome, circle (b) represents density of ncRNA distribution, circle (c) represents density of inverted repeats, circle (d) represents density of direct repeats, circle (e) represents distribution density of genes in reverse arrangement, circle (f) represents distribution density of genes in forward arrangement, and circle (g) represents ostracitoxin GC content. Darker color in each circle indicates higher value. C: Pie chart of genome size in different Tetraodontiformes species, ranked according to genome sizes, ranging from 334.91 Mb (*Arothron firmamentum*) to 864.71 Mb (*O. cubicus*). D: Proportion of different genome elements in *O. cubicus* and related species. E: Proportion of different TEs in *O. cubicus* and related species.

(Figure 2A) and complete mitochondrial genomes from 131 species, including 113 Tetraodontiformes species (Figure 2B). Both datasets consistently placed Lophiiformes as the closest extant group to early-diverging Tetraodontiformes lineages (Figure 2A, B). Mitochondrial phylogeny further revealed close affinity between Tricanthidae, Molidae, Aracanidae, and Ostracidae, supporting a three-clade division: Ostraciidae-Molidae, Tetraodontidae, and Balistidae-Monacanthidae (Figure 2B). Synonymous substitution rate (*Ks*) analysis identified the lowest peak between *O. cubicus* and *M. mola* (*Ks*=0.38), followed by *L. litulon* (0.52), *Tetraodon palembangensis* (0.59), *T. rubripes* and *T. bimaculatus* (0.64), and *Thamnaconus modestus* (0.70), suggesting a particularly close evolutionary relationship between *O. cubicus* and *M. mola* (Figure 2C). Demographic reconstructions of four Tetraodontiformes species revealed diverse population trajectories during the Mid-Pleistocene Transition (MPT) (Figure 2D; Supplementary Figure S6). The “glacial control theory” posits that abrupt sea-level changes during the MPT facilitated intermittent coral reef formation (Daly, 1915), favoring the rapid formation of coral reef ecosystems essential for the survival of Ostraciidae species. This may explain the expansion of the *O. cubicus* population after an initial acclimation phase during the early MPT.

Potential chromosomal evolution in Tetraodontiformes

Ancestral chromosomes within Tetraodontiformes were reconstructed at 300 kb resolution using DESCHRAMBLER, based on chromosome-scale genomes of eight species (one reference, four ingroups, and three outgroups). A total of 600 conserved segments were identified, enabling reconstruction of 24 ancestral chromosomes spanning 663.23 Mb (Figure 3A; Supplementary Table S9). Among ingroups, *O. cubicus* exhibited the fewest chromosomal rearrangements, suggesting a more conserved genome relative to other taxa (Supplementary Table S10). Comparative collinearity analyses between *O. cubicus* and representative species of Eupercaria (*Oryzias latipes*), Molidae (*M. mola*), Tetraodontidae (*T. rubripes*), and Monacanthidae (*Thamnaconus modestus*) revealed multiple lineage-specific rearrangements. In *Oryzias*

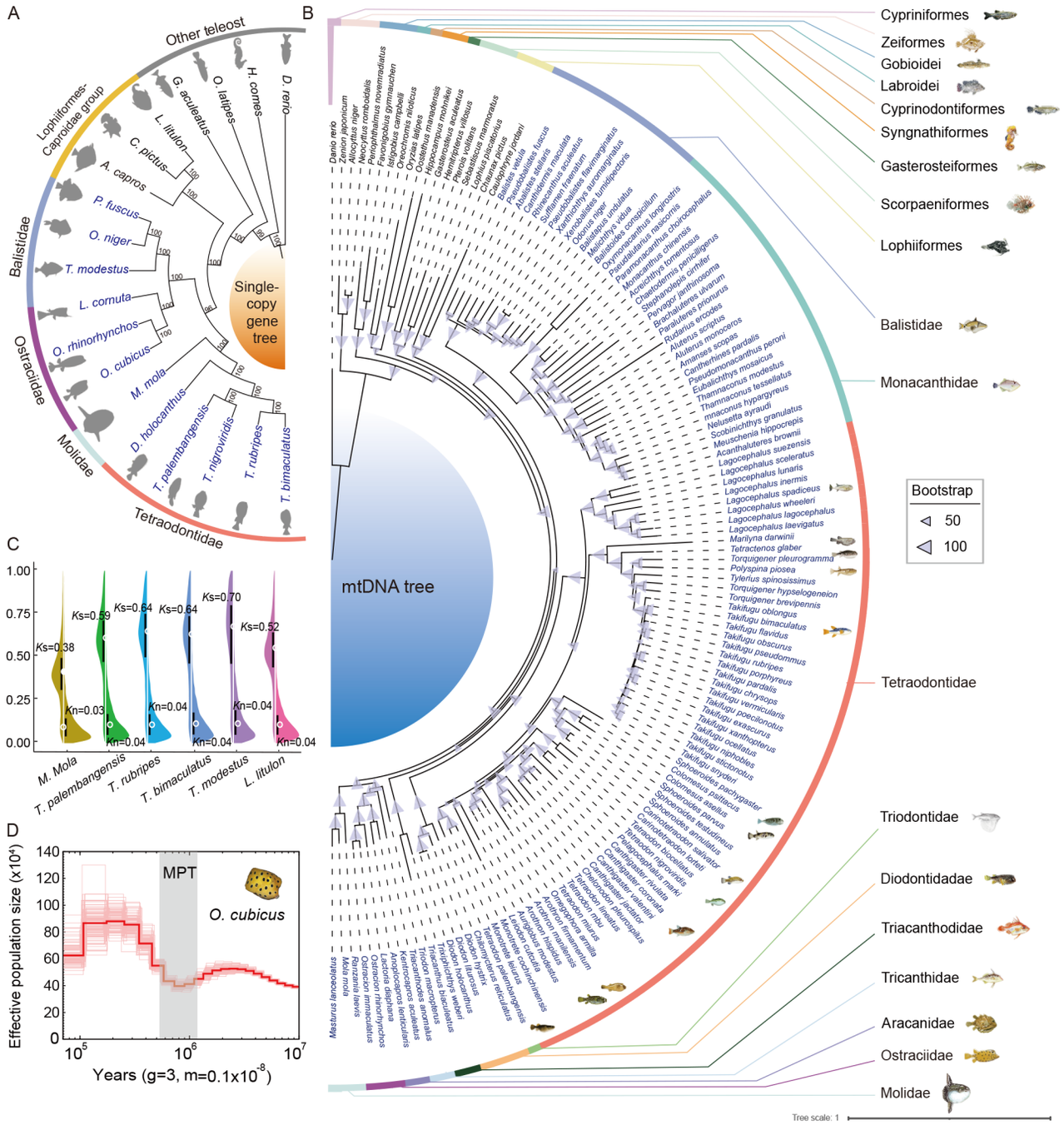

Figure 2 Phylogeny of Tetraodontiformes

A: Maximum-likelihood phylogenetic tree of 12 Tetraodontiformes species (blue) and seven related outgroup species based on single-copy genes. B: Maximum-likelihood phylogenetic tree from complete mitochondrial genomes (mtDNA) of 131 teleost species (including 113 Tetraodontiformes species, marked in blue). Size of purple triangle represents branch bootstrap value. C: Frequency distribution of synonymous and non-synonymous mutation rates of homologous genes in *O. cubicus* and related species. D: Reconstruction of historical population size in *O. cubicus*.

latipes, chromosome 20 was split into chromosomes 20 and 10 in *O. cubicus*, while segments of chromosome 2 underwent translocation and fusion with chromosome 22 to form chromosome 13, with the remaining portion of chromosome 2 contributing to chromosome 6 in *O. cubicus*. Additional rearrangements involved chromosomes 16 and 17 in *O. cubicus* compared to *Oryzias latipes*, *M. mola*, *Thamnaconus modestus*, and *T. rubripes* (Figure 3B, C, D). In *Thamnaconus modestus*, chromosomes 1, 2, 3, and 4 likely originated from the fusion of specific *O. cubicus* chromosomes (Figure 3D). In *T. rubripes*, chromosome 1 resulted from the fusion of *O. cubicus* chromosomes 4 and 18, while chromosome 8 incorporated material from chromosomes 3, 6, and a segment of chromosome 13 (Figure 3D).

Genetic alterations in the *O. cubicus* genome

Molecular clock analyses estimated the divergence between *O. cubicus* and *M. mola* at approximately 94.4 million years ago (Ma), during the Cenomanian-Turonian stage of the Late Cretaceous. Rapid divergence of the Tetraodontidae, Ostraciidae-Molidae, and Balistidae-Monacanthidae lineages occurred over a brief interval between 116.0 Ma and 94.4 Ma

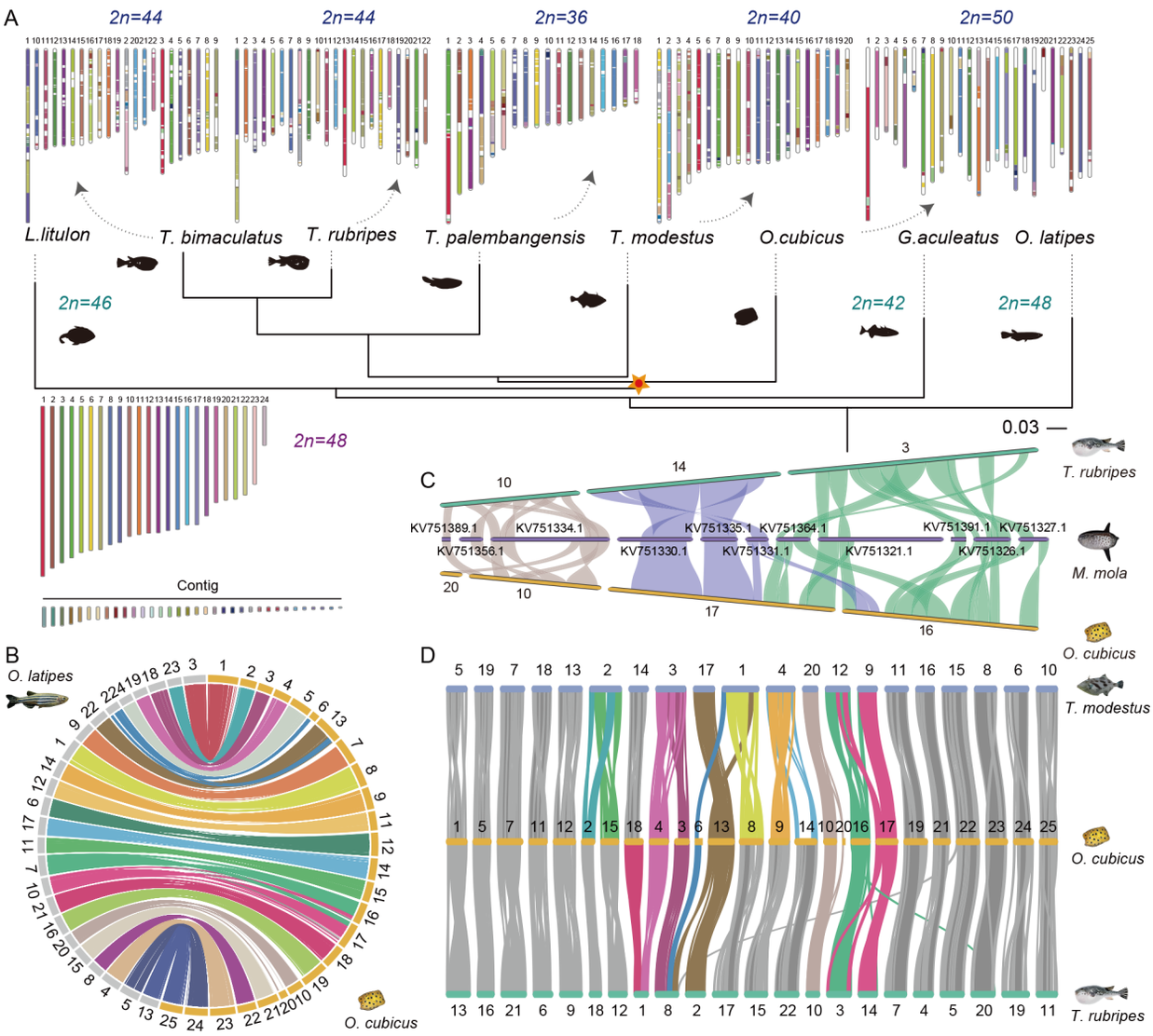

Figure 3 Chromosome evolution of Tetraodontiformes

A: Chromosome evolution of Tetraodontiformes. Star represents divergence of Tetraodontiformes and Lophiiformes. Reference: *O. cubicus*; Ingroup: *T. bimaculatus*, *T. rubripes*, *Tetraodon palembangensis*, and *Thamnaconus modestus*; Outgroup: *L. litulon* (main), *G. aculeatus*, and *Oryzias latipes*. B: Circos plot of *O. cubicus* and *Oryzias latipes* genomes. C: Chromosome collinearity among three species. Compared to *M. mola* and *T. rubripes*, *O. cubicus* had additional rearrangements of chromosomes 16 and 17. D: Collinearity among *O. cubicus*, *T. rubripes*, and *Thamnaconus modestus*.

(Figure 4A). Comparative analyses identified 1 640 expanded gene families and 1 224 contracted gene families in *O. cubicus* (Figure 4A; Supplementary Table S11), with GO and KEGG analysis revealing enrichment in chromosome segregation and cardiac rhythm regulation (Supplementary Figure S7 and Tables S12, S13). In total, 1 313 PSGs and 88 REGs were identified in *O. cubicus* among 10 786 homologous gene pairs, which were confirmed through bidirectional optimal alignment with *M. mola*, *Thamnaconus modestus*, *T. rubripes*, *Oryzias latipes*, and *G. aculeatus* (Supplementary Table S14). Many of these genes were associated with skeletal and scale development, including *bmp7*, *egf7*, *fgf18*, *bmpr2*, *fgfr1a*, *scpp1*, *eomes*, *sost*, and *sct2*. Several REGs, including *lefty2*, *mixl1*, and *hmx2*, were linked to morphological regulation and inner ear development, potentially contributing to carapace formation and balance perception. Functional enrichment of PSGs and REGs further supported roles in ossification, skin morphogenesis, optic vesicle development, and retina homeostasis (Figure 4C, D), implicating these loci in the development of diverse coral reef adaptive phenotypes.

The *SCPP* gene family, essential for the formation of teleost scales, teeth, and bone (Lv et al., 2017), was examined in detail. The *O. cubicus* genome contained 11 *SCPP* genes, including two acidic and nine proline/glutamine-rich (P/Q-rich) variants. Acidic *SCPP* genes regulate bone and dentin mineralization, while P/Q-rich *SCPP* genes are mainly involved in enamel and dentin formation (Lv et al., 2017). Although copy number of *SCPP* genes was conserved across Tetraodontiformes species, gene expression analysis revealed that acidic *SCPP* genes were highly expressed in boxfish carapace tissue compared to the normal skin, while P/Q-rich *SCPP* genes were transcriptionally inactive in both carapace and normal skin (Figure 4E, F). The *HOX* gene family is involved in the regulation of body shape (Cumplido et al., 2024). The Tetraodontiformes order demonstrated a high degree of sequence conservation within the *HoxAb*, *HoxBb*, and *HoxDa* paralogous groups, with no gene copy number

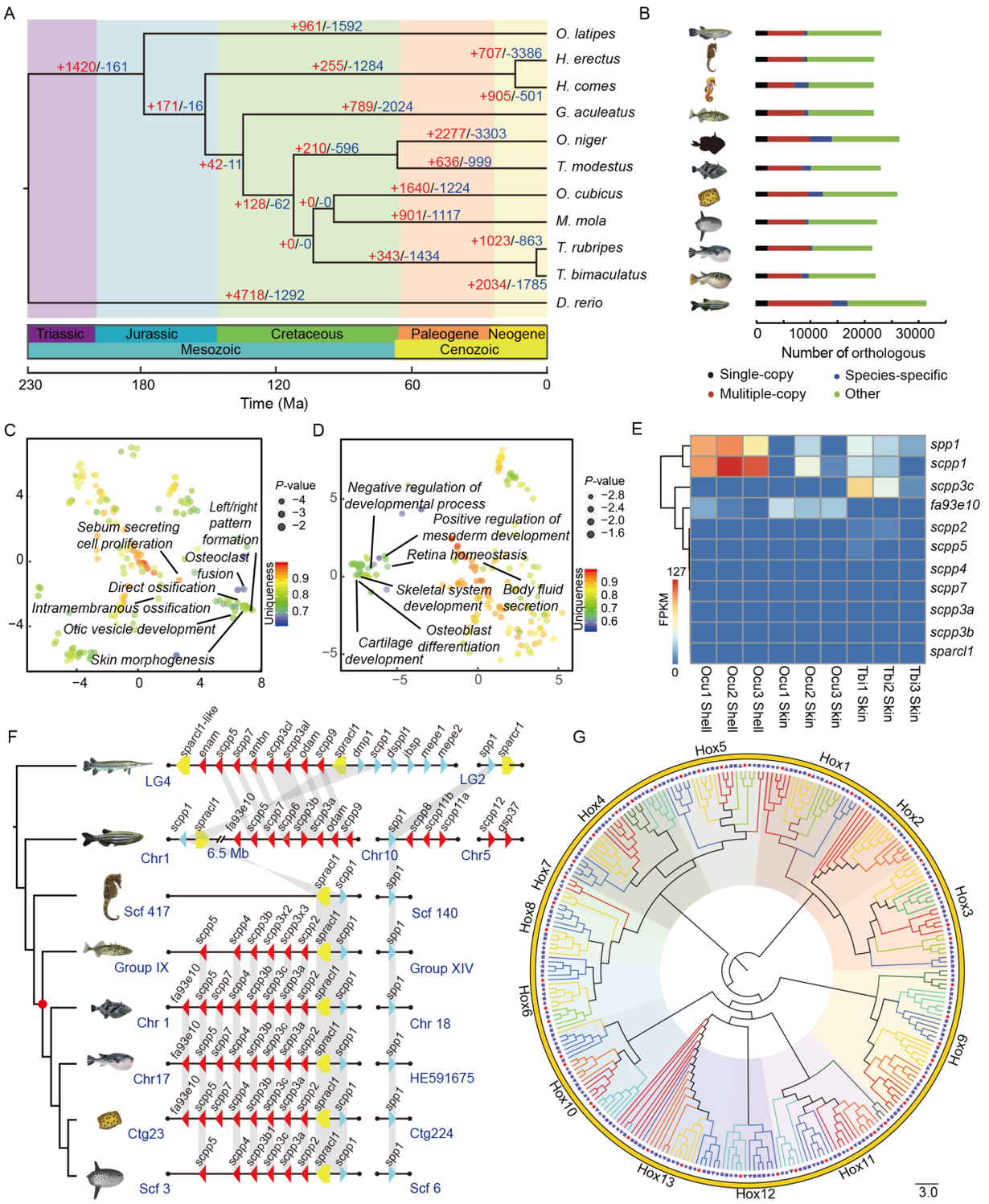

Figure 4 Non-conserved genome elements in *O. cubicus* and related species

A: Divergence time and gene family differentiation of yellow boxfish and related species. Numbers on branches show number of expansion (red)/contraction (blue) genes. B: Types and gene numbers of homologous gene families in yellow boxfish and related species. C: ReviGO enrichment of positively selected genes in yellow boxfish. D: ReviGO enrichment of rapidly evolving genes in yellow boxfish. E: Heatmap of *SCPP* gene family members in yellow boxfish carapace and skin. Gene expression increases from blue to red. Two acidic *SCPP* genes: *spp1* and *scpp1*; Nine proline/glutamine-rich *SCPP* genes: *scpp3c*, *fa93e10*, *scpp2*, *scpp5*, *scpp4*, *scpp7*, *scpp3a*, *scpp3b*, and *sparcl1*. F: Distribution of *SCPP* gene family members in yellow boxfish and related species. G: Phylogenetic relationships among *HOX* gene family members in yellow boxfish and related species.

variations observed in the four representative species (Figure 4G; Supplementary Figure S8). Notably, *O. cubicus* possessed the most complete 48 *Hox* genes among Tetraodontiformes, suggesting a closer evolutionary relationship to ancestral Tetraodontiformes species. Contrary to previous reports of *Hox7* cluster loss in Tetraodontiformes (Ozernyuk & Schepetov, 2022), *O. cubicus*, *M. mola*, and *T. modestus* retained *hoxa7a*, and *O. cubicus* uniquely retained

hoxb7a. The shared presence of *hoxa7a* in *O. cubicus*, *M. mola*, *T. modestus*, and *G. aculeatus* (closely related to Tetraodontiformes) suggests that *hoxa7a* loss of *T. rubripes* may have occurred specifically within the Tetraodontoidei lineage after its divergence from a common ancestor with *O. cubicus* and *M. mola*.

Transcriptomic variation in dermal armor of *O. cubicus*

To identify the molecular basis of coral reef adaptation in *O. cubicus*, morphological and transcriptomic features of the dermal carapace, a typical defensive phenotype, were examined. High-resolution micro-CT of a 3 cm long larval specimen revealed that the carapace adopted a carambola shape, formed by a mosaic of hexagonal bony plates connected through serrated interfaces and supporting protective spines at protruding vertices (Figure 5A). Internally, three pairs of otoliths were located near the brain, consistent with their role in maintaining postural balance in complex reef environments (Supplementary Figure S9A).

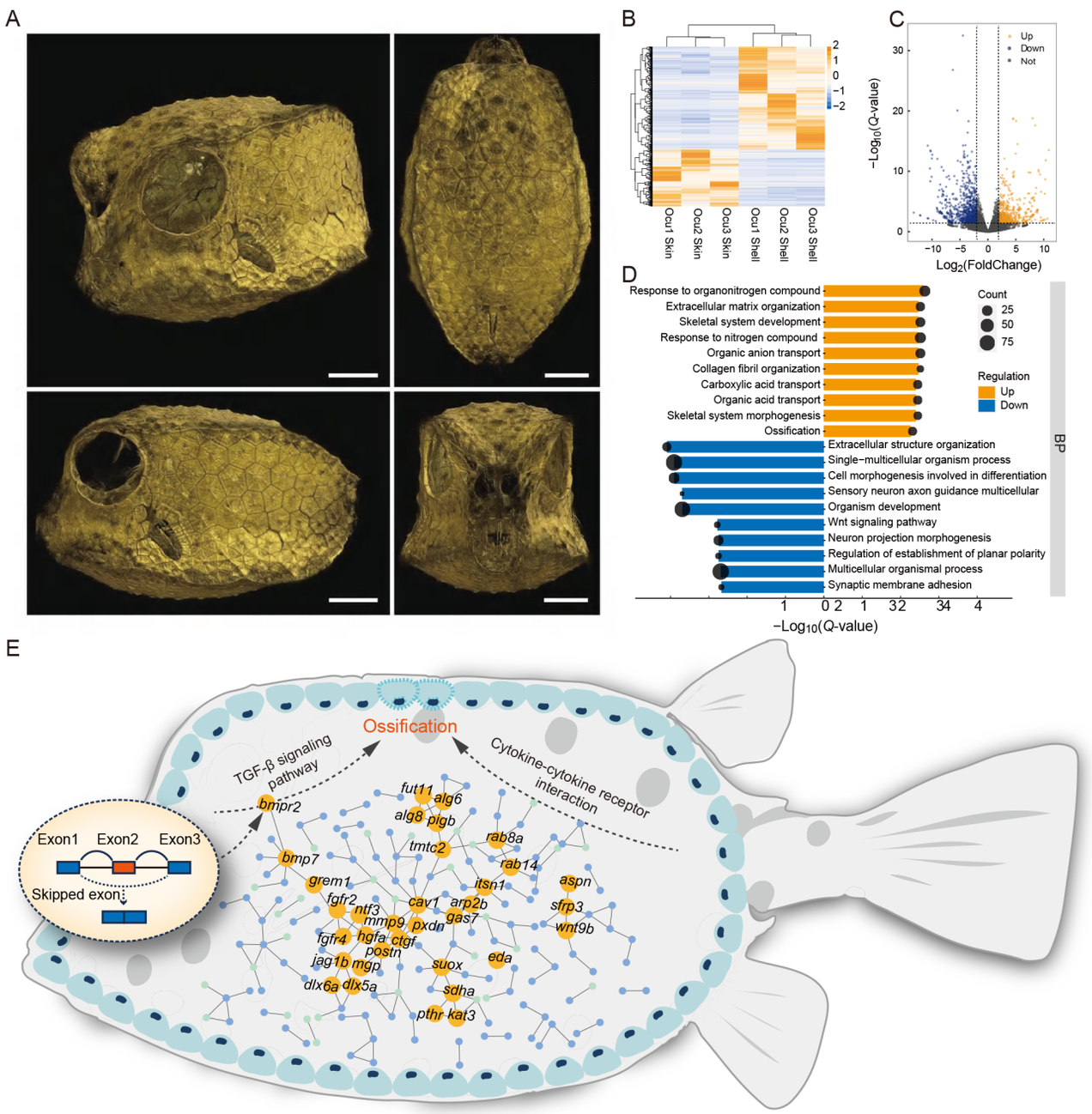

Figure 5 Transcriptional regulation in yellow boxfish carapace formation

A: Micro-CT images of yellow boxfish. Upper left: Perspective view; Upper right: Ventral view; Lower left: Lateral view; Lower right: Anterior view. Scale bars: 5 mm. B: Heatmap of yellow boxfish carapace-specific DEGs. C: Volcano plot of yellow boxfish carapace-specific DEGs. D: Top 10 GO (BP) terms and KEGG pathways enriched in yellow boxfish carapace-specific DEGs. E: Integrated regulation of yellow boxfish carapace formation. Figure shows interaction network of carapace-specific DEGs of yellow boxfish. Each point represents a gene, and lines between points represent interactions between genes. Orange points: Hub genes; Blue points: DEGs; Green points: Differentially spliced DEGs. Osteoblast-like cells (blue cells in picture) ossify to form the boxfish carapace under differential expression and splicing regulation. Ossification is mediated by cytokine (such as TGF-β) signaling.

The morphological results supported transcriptomic sampling of the yellow boxfish carapace and tail skin (Supplementary Figure S9B). Comparative transcriptome profiling identified 600 DEGs between the carapace and tail skin, including 225 down-regulated and 375 up-regulated transcripts in the carapace (Figure 5B, C; Supplementary Table S15). These DEGs showed complex expression patterns between the carapace and tail skin. While several key inducers of bone and cartilage development, such as *bmp7*, *postn2*, and *ogn*, were down-regulated, genes directly associated with ossification, including acidic *SCPP* genes and *bglap2*, were significantly up-regulated. These patterns suggest that carapace formation involves an integrated balance of both inductive and repressive signals, mirroring elements of endoskeletal development while maintaining distinct dermal-specific regulatory features. Furthermore, GO enrichment revealed that up-regulated DEGs in the carapace were enriched in biological processes related to ossification and skeletal system development, whereas down-regulated DEGs were predominantly enriched in cellular morphogenesis (Figure 5D; Supplementary Tables S16, S17). Notably, several genes implicated in skeletal system development and bone morphogenesis, such as *postn* and *bmpr2*, also underwent differential AS (Supplementary Table S18). Skipped exon events occurred most frequently in DAS genes, with overall increases in isoform diversity in the carapace across SE, A5SS, A3SS, and RI events (Supplementary Figure S9C). For example, a *postn* transcript featuring exon 17 exclusion was significantly enriched in the carapace, potentially reflecting isoform-specific functional specialization during carapace formation (Supplementary Figure S9D). The co-expression network of DEGs and DAS genes identified 34 hub regulatory genes, many involved in ossification and skeletal system development pathways. Communication between cytokine (TGF-β) signaling and osteoblast-like cells may be a vital axis in the development of yellow boxfish dermal armor (Figure 5E).

Convergent evolution of *eda* in teleost bony plates

Dermal bony plates have evolved independently across multiple teleost lineages, and the *eda* gene has been implicated as a key regulator of this trait. In *G. aculeatus*, for example, distinct *eda* mutations are correlated with bony development in both marine and freshwater populations (Robertson et al., 2017). While transcriptomic analysis of *O. cubicus* revealed no significant expression difference in *eda* or its downstream receptor *edar* between carapace and skin tissues, a comparative evolutionary analysis of *eda* across 13 teleost species (six with bony plates and seven with no bony plates) uncovered a notable site of convergent evolution at four amino acid mutations (Figure 6). Predicted domain architecture of Eda proteins across all examined species revealed a conserved structure consisting of a transmembrane region and tumor necrosis factor (TNF) domain (Figure 6A). A specific convergence site located just two residues upstream of the predicted α-helical segment within the transmembrane domain emerged as a key signature. At this site, species possessing dermal bony plates —including *O. cubicus*, *Odonus niger*, *Gasterosteus aculeatus*, *Hippocampus erectus*, *H. comes*, and *Atractosteus spatula*—uniformly encoded leucine (L), whereas plate-less species encoded proline (P) or alanine (A) (Figure 6B). The proximity of this substitution to the α-helical anchor suggests a potential influence on membrane insertion or receptor-binding conformation (Figure 6A, B). Three-dimensional (3D) structural modeling and RMSD clustering of the 13 Eda proteins yielded two distinct groups: one exclusively composed of species lacking bony plates (5/5), and the other predominantly comprising plate-bearing taxa (6/8) (Figure 6C). Elevated RMSD values were primarily associated with comparisons between the two structural clusters, implying that bony plate evolution may involve convergent remodeling of Eda tertiary structure. Although overall protein stability predictions showed no significant difference between species with and without bony plates (Figure 6D), substitution of L to P at the convergent site in plate-bearing species resulted in a markedly higher energy cost, indicating structural destabilization. In contrast, the reverse mutations from P or A to L in plate-less species caused relatively minor energy changes (Figure 6E). These findings suggest that L is structurally favored at this site in species with bony plates, and that reverting to P may be selectively disfavored due to its destabilizing effect on Eda protein conformation.

DISCUSSION

Comparative genomic analyses indicated that *O. cubicus* has retained several ancestral features that distinguish it from other members of Tetraodontiformes. At the chromosomal level, *O. cubicus* exhibited markedly fewer chromosomal rearrangements relative to related taxa, suggesting a karyotypic structure more closely aligned with the ancestral condition. At the gene level, the retention of a more complete set of *Hox* cluster members further indicated a closer relationship to the Tetraodontiformes ancestor. Together with its early divergence and phylogenetic position, these characteristics indicate that *Ostracion cubicus* represents an extant crown lineage that remains closer to ancestral Tetraodontiformes species.

Interestingly, despite its overall genomic conservation, the various phenotypic adaptations of *O. cubicus* can be traced to the role of genetic factors. Recent lineage-specific TE bursts and chromosomal rearrangement events likely contributed to regulatory rewiring and genome plasticity. For example, the REG *hmx2* and PSG *hoxb13a*, both implicated in inner ear and hypothalamic development (Feng & Xu, 2010), may support enhanced spatial orientation and sensory integration. The evolution of a rigid dermal carapace encasing the entire body imposes increased energy demands for the high-frequency movement of pectoral fins (Van Wassenbergh et al., 2015). This likely necessitated enhanced oxygen delivery and cardiac output, consistent with the observed expansion of gene families related to heart rate regulation. Furthermore, the observed enrichment in genes governing chromosome segregation supports the hypothesis that karyotypic remodeling through ancestral chromosome fragmentation contributed to *O. cubicus* genome architecture. Compared to normal skin, the yellow boxfish carapace exhibited distinct patterns of gene expression, AS, and functional interactions enriched for pathways involved in ossification and skeletal system development, highlighting transcriptional and post-transcriptional mechanisms underlying adaptation to coral reef environments. These molecular signatures underscore how enhanced structural defense and refined locomotor control have jointly contributed to the ecological success of *O. cubicus* under strong selective pressure.

The evolutionary plasticity of gene regulation is further

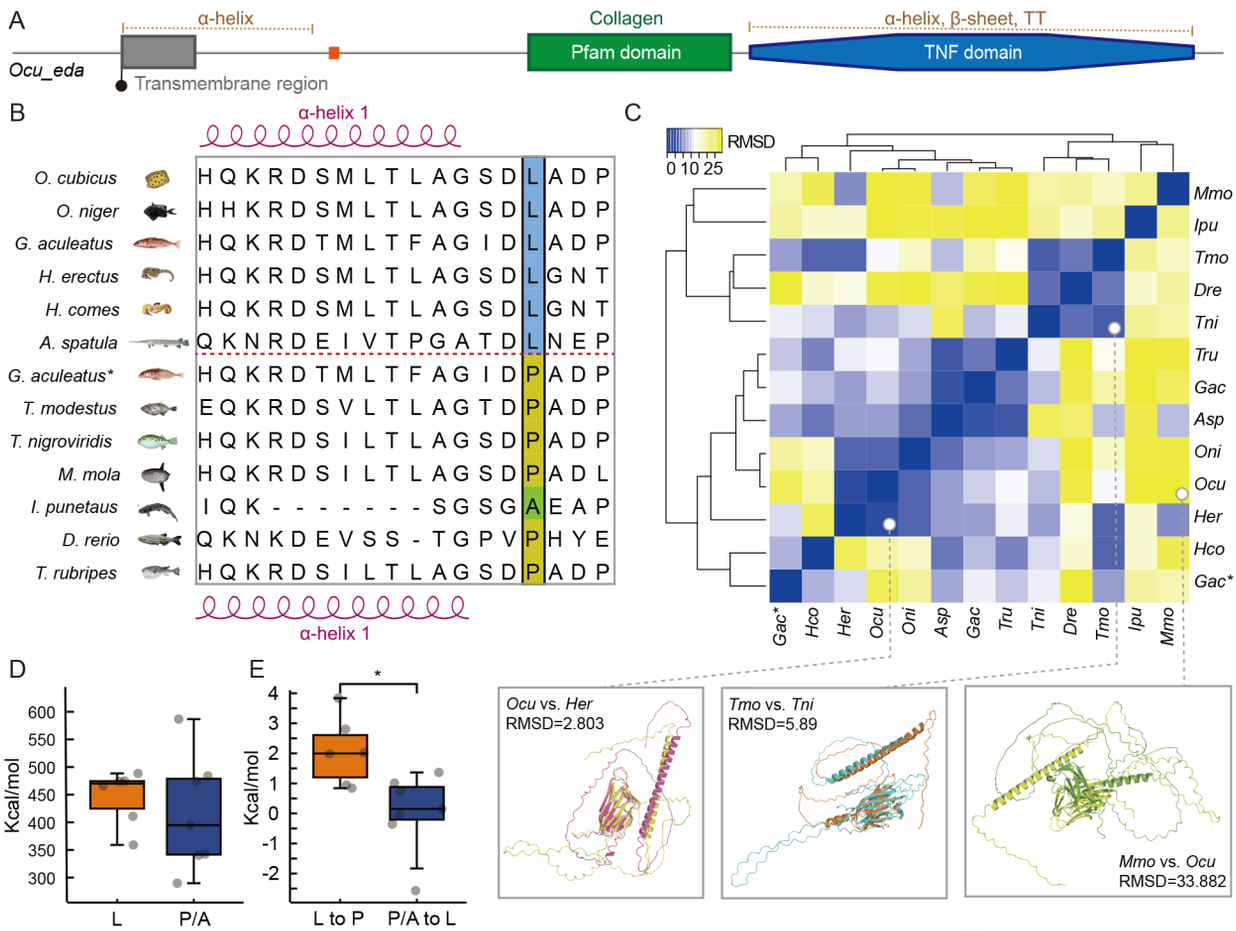

Figure 6 Bony plate convergent evolution in teleost fish

A: Genetically mobile domains of *eda* in yellow boxfish. B: Convergent evolution sites derived from *eda* of six species with bony plates and seven species without bony plates. *G. aculeatus** refers to *G. aculeatus* without bony plate. C: Heatmap of predicted structural similarity of 13 Eda proteins. Root mean square deviation (RMSD) value decreases from yellow to blue, and similarity in 3D structure between two proteins becomes higher with lower RMSD. *G. ac** refers to *G. aculeatus* without bony plate. Three representative 3D structure comparisons are listed under the heatmap. From left to right: *O. cubicus* (with bony plate) and *Hippocampus erectus* (with bony plate), *Thamnaconus modestus* (no bony plate) and *Tetraodon nigroviridis* (no bony plate), and *M. mola* (no bony plate) and *O. cubicus* (with bony plate). D: Predicted overall stability of Eda proteins. E: Eda protein energy differences before and after mutation at convergent evolution sites. *: $P<0.05$. D, E: Each gray point represents predicted protein energy of a species. Orange box: Six species with bony plates; Blue box: Seven species without bony plates.

exemplified in the divergent development of dermal bony plates. Notably, the presence of unarmored caudal skin in *O. cubicus* provides an internal morphological control for studying dermal ossification, enabling within-individual comparisons for divergent evolution study within the order Tetraodontiformes. Several key regulators of bone development appear to play central roles in this transformation: *bmp7* functions as an osteogenic inducer that promotes epithelial ossification (Kim et al., 2012); *bglap2* encodes osteocalcin, which constitutes 1%–2% of bone matrix and binds to apatite and calcium (Xu et al., 2023); *aspn* modulates osteoblast-driven collagen biomineralization by binding to calcium and inhibiting the typical TGF-β/Smad signaling pathway (Zhou et al., 2023); and *germ1* acts as a Bmp2 antagonist, inhibiting Bmp2-mediated osteoblast differentiation (Deng et al., 2021). These findings underscore the integration of genetic selection and transcriptional regulation in shaping dermal armor formation and reveal *O. cubicus* as a valuable model for uncovering the evolutionary origins of skeletal innovation in teleosts.

The dermal armor of the yellow boxfish consists of hard hexagonal scales formed through the spiral accumulation and mineralization of collagen (Yang et al., 2015). This ossified scale-based architecture represents a broader evolutionary pattern observed in multiple teleost lineages, including members of Lepisosteiformes, Syngnathiformes, Tetraodontiformes, and Perciformes, suggesting repeated and independent evolution of bony plates across distantly related Osteichthyes species. Ectodysplasin-A (Eda) is a cytokine involved in epithelial-mesenchymal signaling and skin derivative development (Yang et al., 2022). Comparative studies on freshwater (without bony plate) and marine (with bony plate) dwelling *G. aculeatus* populations demonstrated a strong association between bony plate formation and specific amino acid mutation sites in the *eda* gene (Robertson et al., 2017), including an L to P mutation that coincides with the convergent evolutionary site identified in this study. Translational and structural predictions across a broader taxonomic spectrum revealed that such mutations, particularly those proximal to conserved secondary structural motifs, may have substantial functional consequences (Khan & Vihinen, 2007). Within α-helices, leucine promotes hydrophobic interactions and helps maintain helix stability (Lyu et al., 1991), while proline tends to disrupt helix elongation and its stability (Morgan & Rubenstein, 2013). In Eda proteins from

species with dermal plates, the presence of leucine at the convergent site may enhance transmembrane *α*-helix integrity in *eda* domains through the hydrophobic side chain. In contrast, proline substitutions at this site in species lacking bony plates may restrict protein flexibility and stability. Notably, this pattern of convergence was further supported by analyses of predicted tertiary structures: Eda proteins clustered according to bony plate phenotype rather than phylogenetic affiliation, underscoring the adaptive convergence of structural features at the protein level. This phenomenon mirrors other known cases of molecular convergence, such as the parallel evolution of high-affinity hemoglobin in high-altitude birds (Natarajan et al., 2015). These findings imply that the structural similarity of Eda proteins correlates more strongly with the presence or absence of dermal bony plates than with phylogenetic relatedness. Although individual amino acid changes are often assumed to have limited impact on protein stability, the observed differences in free energy associated with mutations at this site point to more detailed evolutionary trends. Investigating the structural patterns of key genes and proteins may provide deeper insights into the molecular basis of phenotypic convergence, highlighting the need for further exploration of the regulatory mechanisms underlying *eda*-mediated bony plate development in teleosts.

Rapid phenotypic adaptation in response to complex environments can coexist with a conserved genome, a dynamic equilibrium that serves as a fundamental driver of biodiversity and complexity (Pfennig et al., 2010; Promy et al., 2023). However, evolution is complex, involving various mechanisms operating across distinct temporal and molecular scales. In the future, functional validation using gene editing approaches will be applied to verify the functions of several candidate genes involved in *O. cubicus* carapace development, such as *bmp7*, *postn*, *aspn*, *scpp1*, and *eda*. Expanding high-quality genomic datasets from across Tetraodontiformes will further refine phylogenetic resolution, while population-scale analyses will be critical for uncovering the finer mechanisms of coral reef adaptive evolution in *O. cubicus*.

The present study identified an intriguing phenomenon in evolutionary biology: extensive phenotypic diversification can occur even in the context of a highly conserved genome. Through the construction of a high-contiguity chromosome-level genome for *O. cubicus* and comprehensive evolutionary and comparative analyses, *O. cubicus* was identified as a basal lineage within Tetraodontiformes, with a chromosome count resembling the ancestral karyotype, early divergence time, and retention of the largest number of *Hox* genes within the order, hallmarks of genomic conservatism. Environmental shifts during the MPT, particularly glacial expansion and sea-level fluctuations, likely created new reef habitats that shaped *O. cubicus* adaptation. Recent bursts of TEs, lineage-specific chromosomal rearrangements, gene family expansions and contractions, and signatures of positive selection and rapid evolution in phenotype-related genes collectively underpin the genetic foundation of coral reef adaptation in *O. cubicus*. Transcriptomic and splicing analyses of the carapace revealed extensive regulatory modulation of genes involved in ossification and collagen matrix organization, reflecting adaptive mechanisms operating at both the transcriptional and post-transcriptional levels. Cytokine-mediated signaling pathways such as TGF-β appear to play a critical role in carapace formation in yellow boxfish. Furthermore, the conserved amino acid sequence and tertiary structure of *eda* across plate-bearing teleosts suggest a convergent molecular trajectory underlying the evolution of bony plates. Together, these findings offer valuable genomic resources and contribute key evolutionary insights into the genetic basis of dermal carapace development and the broader mechanisms of specialization in coral reef ecosystems.

DATA AVAILABILITY

All data generated or analyzed during this study are included in this published article and its supplementary information files. The *Ostracion cubicus* genome and all raw sequencing data reported in this paper have been deposited in the NCBI database under BioProjectID PRJNA1218305, Genome Sequence Archive (Chen et al., 2021) in the National Genomics Data Center under BioProject PRJCA028019, and Science Data Bank (10.57760/sciencedb.j00139.00164).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

P.X. T.Z., and F.P. conceived and designed the project. S.M.H. and Z.X.Z. are co-first authors and made equal contributions to experimental design, data collection, data analysis, and manuscript writing. J.Y.Y. and Z.J. collected the samples. P.X. and T.Z. serve as co-corresponding authors, providing guidance on the research. Q.M.Q., P.X., and T.Z. revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank all the students and teachers of the Fish Genetics and Breeding Laboratory (FGL), College of Ocean and Earth Sciences, Xiamen University, for their assistance.

REFERENCES

Abrusán G, Grundmann N, Demester L, et al. 2009. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*, **25**(10): 1329–1330.

Álvarez-Carretero S, Tamuri AU, Battini M, et al. 2022. A species-level timeline of mammal evolution integrating phylogenomic data. *Nature*, **602**(7896): 263–267.

Aparicio S, Chapman J, Stupka E, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297**(5585): 1301–1310.

Arcila D, Pyron RA, Tyler JC, et al. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, **82**: 131–145.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**(2): 573–580.

Brůna T, Hoff KJ, Lomsadze A, et al. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, **3**(1): lqaa108.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, **18**(4): 366–368.

Chapman JA, Kirkness EF, Simakov O, et al. 2010. The dynamic genome of *Hydra*. *Nature*, **464**(7288): 592–596.

Chen TT, Chen X, Zhang SS, et al. 2021. The genome sequence archive

family: toward explosive data growth and diverse data types. *Genomics, Proteomics & Bioinformatics*, **19**(4): 578–583.

Cumplido N, Arratia G, Desvignes T, et al. 2024. *Hox* genes control homocercal caudal fin development and evolution. *Science Advances*, **10**(3): eadj5991.

Daly RA. 1915. The glacial-control theory of coral reefs. *Proceedings of the American Academy of Arts and Sciences*, **51**(4): 157–251.

Delano WL. 2002. Pymol: an open-source molecular graphics tool. *CCP4 Newsl Protein Crystallogr*, **40**(1): 82–92.

Delgado J, Radusky LG, Cianferoni D, et al. 2019. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*, **35**(20): 4168–4169.

Deng FM, Zhao JL, Jia W, et al. 2021. Increased hypospadias risk by *GREM1* rs3743104 [G] in the southern Han Chinese population. *Aging*, **13**(10): 13898–13908.

Dudchenko O, Batra SS, Omer AD, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**(6333): 92–95.

Durand NC, Shamim MS, Machol I, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, **3**(1): 95–98.

Edgar RC. 2010. Quality measures for protein alignment benchmarks. *Nucleic Acids Research*, **38**(7): 2145–2153.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**(1): 238.

Feng Y, Xu QL. 2010. Pivotal role of *hmx2* and *hmx3* in zebrafish inner ear and lateral line development. *Developmental Biology*, **339**(2): 507–518.

Flynn JM, Hubley R, Goubert C, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(17): 9451–9457.

Gao Y, Gao Q, Zhang H, et al. 2014. Draft sequencing and analysis of the genome of pufferfish *Takifugu flavidus*. *DNA Research*, **21**(6): 627–637.

Garner SN, Naleway SE, Hosseini MS, et al. 2020. The role of collagen in the dermal armor of the boxfish. *Journal of Materials Research and Technology*, **9**(6): 13825–13841.

Gordon MS, Hove JR, Webb PW, et al. 2000. Boxfishes as unusually well-controlled autonomous underwater vehicles. *Physiological and Biochemical Zoology*, **73**(6): 663–671.

Gremme G, Brendel V, Sparks ME, et al. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, **47**(15): 965–978.

Haas BJ, Salzberg SL, Zhu W, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**(1): R7.

Han MV, Thomas GWC, Lugo-Martinez J, et al. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*, **30**(8): 1987–1997.

Harris R. 2007. Improved pairwise alignment of genomic DNA. Ph. D. dissertation, Pennsylvania State University, Philadelphia.

Holcroft NI. 2005. A molecular analysis of the interrelationships of tetraodontiform fishes (Acanthomorpha: Tetraodontiformes). *Molecular Phylogenetics and Evolution*, **34**(3): 525–544.

Hove JR, O'bryan LM, Gordon MS, et al. 2001. Boxfishes (Teleostei: Ostraciidae) as a model system for fishes swimming with many fins: kinematics. *Journal of Experimental Biology*, **204**(8): 1459–1471.

Huang N, Li H. 2023. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics*, **39**(10): btad595.

Janvier P. 2008. Early jawless vertebrates and cyclostome origins. *Zoological Science*, **25**(10): 1045–1056.

Jones P, Binns D, Chang HY, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**(9): 1236–1240.

Kalmanzon E, Zlotkin E. 2000. An ichthyotoxic protein in the defensive skin secretion of the Red Sea trunkfish *Ostracion cubicus*. *Marine Biology*, **136**(3): 471–476.

Kang S, Kim JH, Jo E, et al. 2020. Chromosomal‐level assembly of *Takifugu obscurus* (Abe, 1949) genome using third‐generation DNA sequencing and Hi‐C analysis. *Molecular Ecology Resources*, **20**(2): 520–530.

Khan S, Vihinen M. 2007. Spectrum of disease-causing mutations in protein secondary structures. *BMC Structural Biology*, **7**: 56.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**(4): 357–360.

Kim HK, Kim JH, Park DS, et al. 2012. Osteogenesis induced by a bone forming peptide from the prodomain region of BMP-7. *Biomaterials*, **33**(29): 7057–7063.

Kim J, Farré M, Auvil L, et al. 2017. Reconstruction and evolutionary history of eutherian chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(27): E5379–E5388.

Koren S, Walenz BP, Berlin K, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research*, **27**(5): 722–736.

Kozlov AM, Darriba D, Flouri T, et al. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, **35**(21): 4453–4455.

Lagesen K, Hallin P, Rødland EA, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35**(9): 3100–3108.

Larkin MA, Blackshields G, Brown NP, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**(21): 2947–2948.

Lemopoulos A, Montoya-Burgos JI. 2021. From scales to armor: scale losses and trunk bony plate gains in ray-finned fishes. *Evolution Letters*, **5**(3): 240–250.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**(5): 955–964.

Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. *In*: Russell DJ. Multiple Sequence Alignment Methods. Totowa: Humana, 155–170.

Lv YY, Kawasaki K, Li J, et al. 2017. A genomic survey of *SCPP* family genes in fishes provides novel insights into the evolution of fish scales. *International Journal of Molecular Sciences*, **18**(11): 2432.

Lyu PC, Sherman JC, Chen A, et al. 1991. Alpha-helix stabilization by natural and unnatural amino acids with alkyl side chains. *Proceedings of the National Academy of Sciences of the United States of America*, **88**(12): 5317–5320.

Manni M, Berkeley MR, Seppey M, et al. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**(10): 4647–4654.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, **27**(6): 764–770.

Marshall NJ, Cortesi F, De Busserolles F, et al. 2019. Colours and colour vision in reef fishes: past, present and future research directions. *Journal of Fish Biology*, **95**(1): 5–38.

Matsuura K. 2015. Taxonomy and systematics of tetraodontiform fishes: a review focusing primarily on progress in the period from 1980 to 2014. *Ichthyological Research*, **62**(1): 72–113.

Morgan AA, Rubenstein E. 2013. Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PLoS One*, **8**(1): e53785.

Natarajan C, Projecto-Garcia J, Moriyama H, et al. 2015. Convergent evolution of hemoglobin function in high-altitude Andean waterfowl involves

limited parallelism at the molecular sequence level. *PLoS Genetics*, **11**(12): e1005681.

Ozernyuk N, Schepetov D. 2022. HOX-gene cluster organization and genome duplications in fishes and mammals: transcript variant distribution along the anterior–posterior axis. *International Journal of Molecular Sciences*, **23**(17): 9990.

Pan HL, Yu H, Ravi V, et al. 2016. The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *Gigascience*, **5**(1): 36.

Pertea M, Kim D, Pertea GM, et al. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, **11**(9): 1650–1667.

Pfennig DW, Wund MA, Snell-Rood EC, et al. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**(8): 459–467.

Porter MM, Ravikumar N, Barthelat F, et al. 2017. 3D-printing and mechanics of bio-inspired articulated and multi-material structures. *Journal of the Mechanical Behavior of Biomedical Materials*, **73**: 114–126.

Promy NT, Newberry M, Gulisija D. 2023. Rapid evolution of phenotypic plasticity in patchy habitats. *Scientific Reports*, **13**(1): 19158.

Qiu HC, Chen LK, Ma XS, et al. 2023. Analysis of heading stability due to interactions between pectoral and caudal fins in robotic boxfish locomotion. *Journal of Bionic Engineering*, **20**(1): 390–405.

Rahner K. 2001. The Trinity. Donceel J, trans. London: Burns & Oates.

Robertson S, Bradley JE, Maccoll ADC. 2017. Eda haplotypes in three-spined stickleback are associated with variation in immune gene expression. *Scientific Reports*, **7**: 42677.

Rozewicki J, Li SL, Amada KM, et al. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research*, **47**(W1): W5–W10.

Santini F, Sorenson L, Alfaro ME. 2013. A new phylogeny of tetraodontiform fishes (Tetraodontiformes, Acanthomorpha) based on 22 loci. *Molecular Phylogenetics and Evolution*, **69**(1): 177–187.

Shannon P, Markiel A, Ozier O, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**(11): 2498–2504.

Shen SH, Park JW, Lu ZX, et al. 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(51): E5593–E5601.

Szklarczyk D, Gable AL, Nastou KC, et al. 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, **49**(D1): D605–D612.

Talla V, Suh A, Kalsoom F, et al. 2017. Rapid increase in genome size as a consequence of transposable element hyperactivity in wood-white (*Leptidea*) butterflies. *Genome Biology and Evolution*, **9**(10): 2491–2505.

Tang HB, Krishnakumar V, Zeng XF, et al. 2024. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta*, **3**(4): e211.

Tarailo‐Graovac M, Chen NS. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, **25**(1): 4.10. 11–14.10. 14.

Tyler JC, Santini F. 2002. Review and reconstructions of the tetraodontiform fishes from the Eocene of Monte Bolca, Italy, with comments on related Tertiary taxa. *Studi e Ricerche sui Giacimenti Terziari di Bolca, Museo Civico di Storia Naturale*, **9**: 47–119.

Van Wassenbergh S, Van Manen K, Marcroft TA, et al. 2015. Boxfish swimming paradox resolved: forces by the flow of water around the body promote manoeuvrability. *Journal of the Royal Society Interface*, **12**(103): 20141146.

Vaser R, Sović I, Nagarajan N, et al. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, **27**(5): 737–746.

Walker BJ, Abeel T, Shea T, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **9**(11): e112963.

Wang Y, Tang H, Debarry JD, et al. 2012. *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**(7): e49–e49.

Waterhouse AM, Procter JB, Martin DMA, et al. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**(9): 1189–1191.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**(suppl_2): W265–W268.

Xu ZH, Yang C, Wu F, et al. 2023. Triple-gene deletion for osteocalcin significantly impairs the alignment of hydroxyapatite crystals and collagen in mice. *Frontiers in Physiology*, **14**: 1136561.

Yamanoue Y, Miya M, Matsuura K, et al. 2008. A new perspective on phylogeny and evolution of tetraodontiform fishes (Pisces: Acanthopterygii) based on whole mitochondrial genome sequences: basal ecological diversification?. *BMC Evolutionary Biology*, **8**: 212.

Yang RH, Mei YL, Jiang YH, et al. 2022. Ectodysplasin A (EDA) signaling: from skin appendage to multiple diseases. *International Journal of Molecular Sciences*, **23**(16): 8911.

Yang TY, Wang ZY, Liu Y, et al. 2019. Population genetics and molecular phylogeography of *Thamnaconus modestus* (Tetraodontiformes, Monachanthidae) in northwestern Pacific inferred from variation of the mtDNA control region. *Aquatic Living Resources*, **32**: 18.

Yang W, Chen IH, Gludovatz B, et al. 2013. Natural flexible dermal armor. *Advanced Materials*, **25**(1): 31–48.

Yang W, Naleway SE, Porter MM, et al. 2015. The armored carapace of the boxfish. *Acta Biomaterialia*, **23**: 1–10.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586–1591.

Zhang R, Li C, Yu MJ, et al. 2021. Chromosome-level genome assembly of the humpback puffer, *Tetraodon palembangensis*. *Gigabyte*, **2021**: gigabyte17.

Zhou GH, Yan XM, Chen ZF, et al. 2023. ASPN synergizes with HAPLN1 to inhibit the osteogenic differentiation of bone marrow mesenchymal stromal cells and extracellular matrix mineralization of osteoblasts. *Orthopaedic Surgery*, **15**(9): 2423–2434.

Zhou ZX, Liu B, Chen BH, et al. 2019. The sequence and de novo assembly of *Takifugu bimaculatus* genome using PacBio and Hi-C technologies. *Scientific Data*, **6**(1): 187.