# The spotted parrotfish genome provides evolutionary insight into the ecological adaptation of a keystone dietary specialist

Yi-Kai Tea1,2#\*, Yulu Zhou3\*, Kyle M. Ewart1\*, Guo Cheng3\*, Kazuhiko Kawasaki4, Joseph D. DiBattista2,5, Simon Y. W. Ho1, Nathan Lo1# & Shaohua Fan3#

1School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia

2Australian Museum Research Institute, Australian Museum, 1 William Street, Sydney, New South Wales 2010, Australia

3State Key Laboratory of Genetic Engineering, Lab for Evolutionary Synthesis, Human Phenome Institute, Zhangjiang Fudan International Innovation Center, School of Life Science, Fudan University, Shanghai, 200438, China

4Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA

5School of Molecular and Life Sciences, Curtin University, Perth, Western Australia 6102, Australia

\*Contributed equally

#Corresponding authors: [yi-kai.tea@sydney.edu.au](mailto:yi-kai.tea@sydney.edu.au), [shaohua\_fan@fudan.edu.cn](mailto:shaohua_fan@fudan.edu.cn), [nathan.lo@sydney.edu.au](mailto:nathan.lo@sydney.edu.au)

**Abstract**

Adaptive radiation as a result of ecological opportunity can have profound effects on the evolutionary outcome of species. On coral reefs, parrotfishes have been considered as one of the most dramatic examples of adaptive radiation unique in their extreme dietary specialisation. Using abrasion-resistant biomineralized teeth, parrotfishes are able to mechanically extract protein-rich micro-photoautotrophs growing in and amongst reef carbonate material. This unique ability to exploit a previously untapped trophic resource is thought to have led to the early diversification of the parrotfishes. In order to better understand the key evolutionary innovations leading to the success of these dietary specialists, we sequenced and analysed the genome of the spotted parrotfish (*Cetoscarus ocellatus*). Our findings reveal significant expansion, selection, and duplication within several gene families responsible for detoxification, including the cytochrome p450 gene family and non-cyp450 carboxylesterases. We find preliminary evidence that the structural mechanism responsible for the extreme hardness and biomineralization of parrotfish teeth may be a result of poly-glutamine expansion in the enamel protein ameloblastin. We also detect expansion and selection for several genes related to pigmentation and sequential hermaphroditism. Together, these results highlight a potentially complex interplay of adaptive radiation and sexual selection operating on coral reef ecosystems.

**KEYWORDS**

Labridae, scarini, coral reef fish, whole genome sequencing, adaptation to dietary toxins

**1 | INTRODUCTION**

Adaptive radiation, where rapid ecological diversification and accelerated speciation of lineages is preceded by colonization and exploitation of novel habitats (Schulter, 2000; Burbink, & Pyron, 2010; Grant, 2013), provides a robust starting point in an effort to answer some of evolutionary biology’s most fundamental questions. That is, why are there so many species, and how is this variation explained? Among aquatic ecosystems, coral reefs are thought to be responsible for the explosive diversification for many marine fish families, particularly during the Paleocene, when tropical scleractinian reefs first emerged (Price et al., 2011; Bellwood et al., 2015). Today, they provide a unique arena in which to explore the interactions between ecological opportunity, key phenotypic innovation, and more recent drivers of evolution such as sexual selection (Hench et al., 2019).

The diversification of the wrasses (family Labridae) has been described as one of the most stunning examples of adaptive radiation on coral reefs (Streeman et al., 2002; Westneat et al., 2005). The contemporary success of this major teleostean radiation can be attributed to their ability to exploit nearly every trophic guild (Evans et al., 2019; Huertas & Bellwood, 2020). However, it is in the parrotfishes that these specialist ecomorphological adaptations are most dramatically displayed. Parrotfishes are unique among coral reef fishes in both their dietary specialisation and associated ecomorphology. Unlike other labrids, their dentition consists of abrasion-resistant, biomineralized teeth fused into plates that are secondarily strengthened by a coating of enamel (Marcus et al., 2017). They use this beak-like dentition to mechanically extract micro-photoautotrophs, primarily filamentous cyanobacteria and crustose coralline algae (Nicholson & Clements 2023), growing on the surface of or within recently dead coral. They do so by first rasping and excavating at calcareous reef matrices, then pulverizing the material using their pharyngeal jaws (Clements & Choat, 2018; Nicholson & Clements, 2020). This ingested material is passed through the digestive system and ejected as sediment, making parrotfishes keystone ecosystem engineers responsible for reef bioerosion, sediment generation and transport, and the general maintenance of coral reefs (Bonaldo et al., 2014).

In addition to their mechanical inaccessibility, cyanobacterial filaments are thought to be defended by allelochemicals and other xenobiotic toxins, though the exact extent of cyanobacterial toxicity is currently unclear (Nicholson & Clements, 2023). Nonetheless, the ability for parrotfishes to subsist on a diet that is both physically and potentially biochemically inaccessible to most other organisms is remarkable among coral reef fishes. In most terrestrial systems, detoxification is the primary mechanism in which animals cope with exogenous dietary toxins, predominantly through cytochrome p450 and/or epoxide hydrolase mediated pathways (Klaassen & Watskin, 2010; Skinner et al., 2011; Moldonado et al., 2016; Nicholson & Clements, 2023). While tolerance to dietary allelochemicals has been extensively researched in terrestrial systems (Nelson, 2009), comparatively little is known in marine systems.

Aside from their unique dietary specialisation, the biology of parrotfishes is otherwise similar to most other labrid fishes. Parrotfishes are sexually dimorphic, sequential protogynous hermaphrodites, capable of sex changing from female to male (Streeman et al., 2002; Lamm et al., 2015). Males are typically the larger and more colourful sex, and in many cases are adorned with secondary sexual characteristics such as larger and more elaborate fins. In addition to pronounced sexual dichromatism and diandry, more than 75% of parrotfish species also exhibit territorial behaviour, forming large intraspecific harems with females outnumbering males. These traits are strongly correlated with sexual selection, suggesting that reproductive isolation and sexual selection may be factors contributing to the diversification of parrotfishes (Warner & Robertson, 1978; Streeman et al., 2002; Hench et al., 2019).

The remarkable biology and morphology of the parrotfishes has attracted considerable research attention in recent years, including the development of new genomic frameworks in which to explore their evolutionary histories (Kazancioglu et al., 2009). Given the range of features that have enabled their evolutionary success, parrotfishes serve as an ideal candidate for exploring the complex interplay of selective forces operating on coral reefs. In order to better understand the evolutionary drivers responsible for the success of this lineage, and more broadly explore the effects of adaptive radiation and ecological opportunity in coral reef ecosystems, we sequenced an analysed the genome of the spotted parrotfish (*Cetoscarus ocellatus*), a species notable among parrotfishes for having a diet comprising of both cyanobacteria mats and live coral (Nicholson & Clements, 2020). Specifically, we looked for genomic evidence underlying two key evolutionary question concerning parrotfish evolution: (i) are there any metabolic, dietary, or morphological innovations associated with ecological specialisation; and (ii) is sexual selection operating in tandem with adaptive radiation?

Our findings reveal significant expansion, selection, and duplication within several gene families responsible for detoxification, including the cytochrome p450 gene family and other non-cyp450 carboxylesterases, suggestive of the spotted parrotfish’s ability to neutralize dietary toxins potentially associated with cyanobacterial filaments. We find preliminary evidence that the structural mechanism responsible for the extreme hardness and biomineralization of parrotfish teeth may be a result of a previously unreported poly-glutamine expansion in the enamel protein ameloblastin. We also reveal the expansion and selection for several genes related to pigmentation and sex change. Together, these results highlight the complex interplay of adaptive radiation and sexual selection operating on coral reef ecosystems.

**2 | MATERIALS AND METHODS**

**2.1 | DNA extraction and library preparation**

We obtained 50 mg of flash-frozen muscle tissues of the spotted parrotfish (**Figure 1**) from the Australian Museum Research Institute’s tissue collection (specimen voucher AMS I.49509-002). Tissue samples were crushed using a sterile pestle in a 1.5 mL Eppendorf tube containing 600 µL SDS Buffer and 20 µL Proteinase K (Roche, Switzerland) and incubated at 56 °C overnight. Following tissue lysis, 5 µL of RNAse A (20 mg/mL) was added and incubated at 37 °C for 1 h. We extracted DNA using the phenol/chloroform method and checked the concentration using Qubit fluorometric quantitation (ThermoFisher, USA). DNA quality was visualized using 1% agarose gel electrophoresis.

Genomic libraries were prepared according to the manufacturer’s instructions. The single-molecule real-time sequencing bell (SMRT Bell) library was prepared using a PacBio DNA Template Prep Kit 1.0 (Pacific Biosciences). DNA quantitation was performed using a 2100 Bioanalyzer (Agilent Technologies). The SMRT Bell-polymerase complex was constructed using a PacBio Binding Kit 2.0 (Pacific Biosciences), based on the manufacturer’s instructions. The library was loaded onto SMRT cells (Sequel SMRT Cell 1M v2, Pacific Biosciences) and sequenced using Sequel Sequencing Kit 2.1.   
  
Transposase Enzyme-Linked Long-read Sequencing (TELL-seqTM; [Chen et al., 2020]) data (i.e. long-range linked-read data) were generated by the Australian Genome Research Facility. TELL-seq libraries were constructed using a TELL-seq WGS Library Prep Kit (Universal Sequencing Technology). DNA extracted from the spotted parrotfish was incubated with ~8 million TELL beads for barcoding in a 0.2 mL PCR tube according to the manufacturer’s protocol (performed by AGRF). Each TELL bead comprises 50,000 copies of at least one barcode sequence conjugated to its surface. Following barcoding, an eight-cycle amplification step was carried out in order to generate libraries for sequencing.  
  
Hi-C library preparation was performed using the Arima Hi-C Plus kit (Arima, USA) following the manufacturer’s protocol for large animal tissue. Briefly, up to 50 mg of fresh muscle tissue was flashed frozen and pulverized using liquid nitrogen. A crosslinking buffer containing 1% formaldehyde was added to the pulverized tissue to crosslink DNA. 500 ng of DNA was used for Hi-C library preparation. The final library preparation for sequencing was performed using the KAPA Hyper Prep kit following the protocol detailed in the Arima Hi-C Plus kit. The libraries were quantified using a Qubit fluorometer (ThermoFisher, USA) and Bioanalyzer (Agilent, USA).  
  
In order to estimate the number of available Hi-C reads in the library, the library was first sequenced on the Illumina Miseq platform (Illumina, USA) using a Miseq V2 Nano flow cell with 2×150 bp specification. The sequencing data were then processed using qc3C (DeMaere & Darling, 2021). The qc3C software assessed the Hi-C library quality by calculating proximity ligation events which create k-mers that would not naturally occur in the sample. Based on the qc3C result, we estimated the sequencing data needed for the library. Sequencing was carried out on the Novaseq platform (Illumina, USA) using Novaseq S4 flow cell 2×150 bp at Novogene (USA)

**2.2 | Genome Assembly**

An initial de novo assembly of the PacBio long-read data was performed using Canu v.2.1.1 (Koren et al., 2017) with default settings and an estimated genome size of 1.4 Gb (based on published labrid genomes (Mattingsdal et al., 2018; Liu et al., 2021). Putative regions of duplications (i.e., haplotigs) and heterozygous overlaps were identified and removed based on sequence similarity and read depth using purge\_dups v1.2.5 (Guan et al., 2020). We used minimap2 v2.18 (Li, 2018) to align raw PacBio reads to the draft assembly. The assembled genome was then polished using the PacBio reads with the program gcpp (https://github.com/PacificBiosciences/gcpp), implementing the ‘arrow’ algorithm, with the minimum coverage cutoff required for variant calling set to 20, the maximum coverage set to 80, and the minimum confidence for a variant call set to 20.  
  
TELL-seq linked reads were used to scaffold the draft de novo long-read assembly and improve its contiguity. Raw TELL-seq reads were demultiplexed and processed using Tell-Read (Chen et al., 2020). To ensure that TELL-seq reads were compatible with downstream 10X programs, barcodes were downsized to 4,000,000 barcodes, compressed, and renamed using ust10x (provided with the Tell-Read software). Barcodes were further processed and trimmed, and sequences interleaved with Long Ranger basic v2.2.2. For all Long Ranger runs, the barcode whitelist was replaced to match TELL-seq barcodes. Putative scaffolding graphs were generated using ARCS v1.2 (Yeo et al., 2018) and joined using LINKS v1.8.7 (Warren et al., 2015) with parameters ‘arcs-make arcs c=3 a=0.9 l=3 z=500 e=70000 m=10-250’. The optimal settings were identified through a parameter sweep with ARKS/LINKS.  
  
Hi-C reads were aligned to the ARCS/LINKS-scaffolded draft assembly with bwa mem v0.7.17 (Li & Durbin, 2009) using the '-SP5M’ flags (**Table S5**). PCR duplicates were removed using samtools v1.6 (Danecek et al., 2021). The resulting alignment was converted to a ‘bed’ file using bedtools v2.29.2 (Quinlan & Hall, 2010). Contig lengths were calculated using samtools. Scaffolding was performed using the program SALSA v2.3 (Ghurye et al., 2019), with the restriction sites set as ‘GATC’ and ‘GANTC’. A Hi-C contact map was generated by converting SALSA2 outputs with juicer-tools pre v1.22.1 and visualized using Juicebox v2.04.06 (**Figure S1**; Durand et al., 2016).  
  
Polishing with long and linked-read data was conducted on the Hi-C scaffolded assembly. A second round of polishing using the long-read PacBio data was performed as described above (i.e., implementing the ‘arrow’ algorithm). Two further rounds of polishing were conducted with the TELL-seq linked reads. The reads were aligned to the draft assembly with Long Ranger align v2.2.2, then scaffolds were split and variants were called in each scaffold using FreeBayes v0.9.21 (Garrison & Marth, 2012) with default settings. Variants were concatenated, normalized, and filtered to retain only homozygous ALT and non-REF sites. Consensus sequences were generated with bcftools v1.9.  
  
To evaluate the contiguity of the spotted parrotfish genome (i.e. assessing the genome and scaffold lengths), we analyzed the assembled scaffolds with QUAST v4.3 (Gurevich et al., 2013). The evaluate the completeness of the genome, we implemented a Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0 analysis with default parameters, based on a core set of genes of the Actinopterygii database.

**2.3 | Genome annotation and repeat analysis**

Genome annotation was performed using Fgenesh++ v7.2.2 (Salamov & Solovyev, 2000; Solovyev et al., 2006). A repeat database was first generated using RepeatModeler v1.0.8 (Flynn et al., 2020). Repeats were subsequently masked in the genome assembly using the program RepeatMasker v4.0.6 (Smit et al., 2015), implementing the ‘-nolow’ option. The masked and unmasked genome assemblies were used as inputs for Fgenesh++. Annotation using Fgenesh++ was performed with optimized gene-finding parameters trained on the genome of the three-spined stickleback (*Gasterosteus aculeatus*) and with non-mammalian general pipeline parameters. We utilized the NCBI animal protein database, curated by Softberry (Solovyev et al., 2006), for the “prot\_map” homology-based predictions. For genomic regions that did not contain “prot\_map” predictions, *ab initio* predictions were performed. The *ab initio* predictions were then used in a BLAST search against the NCBI database, and predictions without any matches were excluded from subsequent analyses.  
  
Homologous snRNAs and miRNAs were predicted by searching the Rfam database (Kalyari et al., 2018) using Infernal v1.1.2 (Nawrocki & Eddy, 2013) with settings cmscan -Z 2735.647178 --cut\_ga --rfam --nohmmonly. tRNAs and rRNAs were predicted using tRNAscan-SE v2.0.8 (Lowe & Eddy, 1997) and RNAmmer v1.2 (Lagesen et al., 2007), respectively, with default settings.  
  
**2.4 | Orthology assignment, gene family analysis, and divergence dating**

The coding sequences of 14 species, including coelacanth (*Latimeria chalumnae*) (Amemiya et al., 2013), human (*Homo sapiens*) (Lander et al., 2001), spotted gar (*Lepisosteus oculatus*) (Braasch et al., 2016), zebrafish (*Danio rerio*) (Howe et al., 2013), Nile tilapia (*Oreochromis niloticus*) (Brawand et al., 2014), Japanese medaka (*Oryzias latipes*) (Kasahara et al., 2007), three-spined stickleback (*Gasterosteus aculeatus*) (Jones et al., 2012), Japanese pufferfish (*Takifugu rubripes*) (Aparicio et al., 2002), oceanic sunfish (*Mola mola*) (Pan et al., 2016), and gilthead seabream (*Sparus aurata*) (Pérez-Sánchez et al., 2019) were downloaded from the Ensembl database v107 (Cunningham et al., 2022). We downloaded the coding sequences of the copperband butterflyfish (*Chelmon rostratus*) (Fan et al., 2020), New Zealand spotty (*Notolabrus celidotus*), ballan wrasse (*Labrus bergylta*) (Lie et al., 2018) and humphead Maori wrasse (*Cheilinus undulatus*) (Liu et al., 2021) from the NCBI RefSeq database (O’Leary et al., 2016). The proteome and coding sequences of the corkwing wrasse (*Symphodus melops*) were downloaded from <https://figshare.com/articles/dataset/maker_aed-1_0_pfam_eggnog_gff/5589997> and <https://figshare.com/articles/dataset/Symphodus_melops_fasta/5590003>.  
  
We extracted the longest isoform per gene using get\_longest\_transcript.py. We inspected the spotted parrotfish proteome and ensured that there were no alternative isoforms. Orthologous proteins among the 16 species were identified using OrthoFinder v2.5.4 with -S diamond -M msa -T raxml -t 8 -a 8 (Emms & Kelly, 2019). Orthogroups were annotated according to the longest protein sequence in each orthogroup by eggnog-mapper with default settings.  
  
We tested for expansions of gene families under a birth-death model using CAFE v5.0 (Mendes et al., 2020), using a dated species tree (see below) and gene counts from OrthoFinder as inputs. Highly variable gene families were excluded to avoid errors in parameter predictions. We used cafetutorial\_clade\_and\_size\_filter.py to filter out gene families that contained any species with over 100 copies of that gene.

A time-calibrated tree was constructed with r8s (<http://loco.biosci.arizona.edu/r8s/>) using the SpeciesTree\_rooted.txt generated by OrthoFinder as input. Node calibrations used were divergence times between two species from Timetree (<http://timetree.org/home>).  
  
**2.5 | Inference of genes that have undergone positive selection**

To prepare and align sequences from the 16 target species for positive selection analysis, we first used MACSE v1.2 (Ranwez et al., 2011) to remove stop codons and correct for potential frameshift mutations in the coding sequences that were potentially introduced during alignment trimming. The coding sequences were aligned using PRANK v.170427 (Löytynoja & Goldman, 2010) using codon-aware mode. We removed the poorly aligned regions using BMGE v1.12 (Criscuolo & Garibaldo, 2010) with codon-aware mode.  
  
We used the optimized branch-site model (Yang & Nielsen, 1998; Zhang et al., 2005) in CodeML (Yang et al., 2007) to detect signatures of positive selection in the protein-coding genes of the spotted parrotfish. In the branch-site model, the null hypothesis assumes that all branches or all branches and codons evolve neutrally (ω ≈ 0), while the alternative model allows a subset of sites in a specific lineage to have a different ω ratio or accelerated nonsynonymous substitution rate (Yang, 1998; Yang & Nielsen, 1998; Zhang et al., 2005). Based on the results of the branch-site models, we calculated the P-values based on likelihood ratio tests (LRT) and corrected the P-values using an FDR approach (<https://github.com/StoreyLab/qvalue>). Genes with an FDR-adjusted P-value of < 0.2 were identified as undergoing positive selection (Tong et al., 2021).  
  
To explore the functional impact of the positively selected genes or expanded gene families in the spotted parrotfish genome, we ran a protein sequence annotation using eggNOG-mapper (Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021) and conducted a gene-based gene ontology (GO) or a Kyoto Encyclopedia of Genes and Genome Pathways (KEGG) signal pathway enrichment test using clusterProfiler v4.2.2 (Wu et al., 2021) in RStudio with the genes of zebrafish as a representation of each gene family.   
  
**2.6 | Identification of *scpp* and *ambn* gene families**

In order to investigate the evolution of the parrotfish’s unique dentition (**Figure 1c­–d**), we looked at the evolution of key teeth and enamel forming genes in vertebrates. *Scpp* genes were manually identified in the spotted parrotfish genome as well as the following teleost genomes: humphead Maori wrasse, New Zealand spotty, gilthead seabream, Japanese pufferfish, and the three-spined stickleback. We looked for *scpp* genes using TBLASTN (Altschul et al., 1990) with default settings using amino acid sequences encoded by previously known *scpp* genes of teleosts and non-teleost actinopterygians. *scpp* genes were also identified using syntenic analysis. Previous studies have shown that the arrangement of zebrafish *scpp* genes can be traced back to four *scpp* gene clusters that arose from two primordial clusters through teleost genome duplication (Braasch et al., 2016). In the present study, we first searched for genomic regions syntenic to these four *scpp* gene clusters using TBLASTN with amino acid sequences encoded by genes located adjacent to the four *scpp* gene clusters. For *scpp* genes from the humphead Maori wrasse, new zealand spotty, gilthead seabream, Japanese pufferfish, and three-spined stickleback, we searched RNA-seq intron-spanning reads in these genomic regions, which are available in the gene database of GenBank. *scpp* genes of spotted parrotfish were subsequently identified with TBLASTN using these *scpp* gene sequences.

We investigated exon 4 of the ameloblastin gene (*ambn*) in the spotted parrotfish and five other labrid species (ballan wrasse [*Labrus bergylta*; GCF\_900080235.1, cunner [*Tautogolabrus adspersus;* GCA\_020745685.1], corkwing wrasse [*Symphodus melops*; GCA\_947650265.1, blue-head wrasse [*Thalassoma bifasciatum;* GCA\_008086565.1], and the sheepshead wrasse [*Semicossyphus pulcher;* GCA\_022749685.1]),identified with TBLASTN in the whole-genome shotgun contigs (wgs) database of GenBank.

Amino acid sequences encoded by *ambn* were manually aligned. Based on this amino acid sequence alignment, a gene tree was constructed using MEGAX by the maximum likelihood method (Nei & Kumar, 2000; Kumar et al., 2018). For constructing the tree, the best-fit amino acid substitution model (JTT+I model) was calculated also using MEGAX. All gaps were excluded from this analysis, and the reliability of the interior branches was calculated by 500 bootstrap samples (Nei & Kumar, 2000).

**2.7 | Identification of pigmentation-related genes**

We collected a gene list associated with pigmentation. This list consists of 996 genes involved in pigmentation phenotypes that was complied with manual curation of genes annotated in HPO (<https://hpo.jax.org/app/>), Color Genes (<http://www.ifpcs.org/colorgenes/>) and genes reported in previous studies and combined with a cross-species gene list of 650 genes for pigmentation biology in a review study (Baxter et al., 2018). In order to determine whether corresponding pigmentation-related genes in fish were available, we used the BLAST function in NCBI database, resulting in the retention of 928 pigmentation-related genes (PRG). Gene family expansions/contractions for these pigmentation-related genes based on the CAFÉ analyses were compared between the 16 analysed species.  
  
**2.8 | Identification of detoxification-related genes**

We aligned the protein sequences of detoxification-related genes using MAFFT v7.110 (Katoh & Standley, 2013) and trimmed the poorly aligned regions using trimAl v1.4.rev22 (Capella-Gutiérrez et al., 2009) with default parameters. The substitution model of each gene family was inferred using prottest-3.4.2 with default parameters. We inferred maximum-likelihood gene trees with RAxML-NG v1.0.3 (Kozlov et al., 2019) with 1000 bootstraps. The gene trees were visualized using iTOL (Letunic & Bork, 2021).  
  
To detect syntenic relationships around the *ces2b* locus, collinear blocks among the 16 species in the present study were identified using MCScanX (Python version) (Tang et al., 2008). Specifically, we first obtained the longest isoform per gene sequences from the proteome of the sixteen different species using an in-house Python script (get\_longest\_transript.py). We converted the GFF to a BED file using the jcvi.formats.gff script from MCScanX. The proteome sequences between two species were aligned using jcvi.compara.catalog with the ‘--cscore=.99’ parameter. Syntenic genes around *ces2b* loci were then extracted and plotted with the jcvi.graphics.synteny script.

**3 | RESULTS**

**3.1 | Genomic landscape and orthology assignment**

Our assembly methods produced a 1.37 Gb genome assembled in 727 contigs, with an N50 of 4.98 Mb (**Table S2**). Our automated annotation pipeline annotated 33,835 genes (excluding *ab initio* annotations with no BLAST hits), including 98.5% of the 3,640 benchmarking universal single-copy orthologs (BUSCOs) (**Table S3**), the highest value(Seppey et al., 2019) for any published teleost genome and comparable to that of the human assembly. Repetitive sequences account for 52% of the genome, with DNA transposons (Class II, 14.08% accounting for 192 Mb) more abundant than retroelements (Class I, 5.36% accounting for 73 Mb) (**Figure 2a–b; Table S4**). Among other available labrid genomes, this proportion of repetitive sequence is comparable only to the humphead Maori wrasse (*Cheilinus undulatus*; DNA transposons 21.25%, accounting for 249 Mb; retroelements 3.82%, accounting for 45 Mb; **Figure 2c**). The proportion of DNA transposons and retroelements for two other labrid species, the ballan (*Labrus bergylta*) and corkwing (*Symphodus melops*) wrasses, were significantly lower than in the spotted parrotfish and humphead Maori wrasse (DNA transposons: 7.83% and 3.90%, accounting for 63 Mb and 24 Mb; retroelements: 8.74% and 3.70% accounting for 70 Mb and 23 Mb in ballan and corkwing wrasses, respectively; **Figure 2d–e**).

We detected two waves of transposable element (TE) expansions (**Figure 2b**), characterized largely by DNA transposons, in the genome of the spotted parrotfish, and one wave of TE expansion in the humphead Maori wrasse genome (**Figure 2c**). Our findings suggest that TE expansion resulting in genome expansion is likely to have occurred once in the common ancestor of the spotted parrotfish and humphead Maori wrasse, but not within the Labrini (no significant TE expansion was detected in the ballan and corkwing wrasse genomes).

When combined with the other 15 genomes, a total of 374,640 annotated genes (97%) were assigned into 21,832 orthogroups; 7,641 orthogroups (35%) had all species present and 1,882 were single-copy orthogroups (8.62%; **Figure S2**).

**3.2 | Gene expansions and contractions**

We identified 2,190 and 502 gene families with signals of expansion or contraction, respectively, in the spotted parrotfish genome (**Figure 3a**), including significant expansion and significant enrichment (P < 0.05; **Figure 3b**) of eight gene families (*ces2*, *gsta*, *adh1*, *fmo5*, *cyp3a65*, *nat10l*, *cbr1*,and *gstm,* **Figure S3**) that were associated with xenobiotic metabolism.

Our analyses also revealed duplications of *ces2* in the spotted parrotfish and Japanese pufferfish (*Takifugu rubripes*) genomes, resulting in 12 copies of each respectively (**Figure 3c**). These genes encode for several carboxylesterases that are highly expressed in the liver and are responsible for the transformation and metabolism of xenobiotic compounds into more harmless metabolites (Tseng et al., 2005). Specifically, the duplicated copies of *ces2* in the spotted parrotfish and Japanese pufferfish genomes belong to the *ces2b* lineage and arose independently through species-specific duplications (**Figure 3c**). Synteny analyses of the *ces2b* locus revealed that the neighbouring genes of *ces2b* are collinear in the genomes of the spotted parrotfish and the 12 other teleosts examined here, suggesting a conserved synteny at this locus during their evolutionary histories. These results also indicate that the extra copies of *ces2b* in the spotted parrotfish genome and the Japanese pufferfish genome likely arose through tandem duplication events (**Figure 2d,e–E** and **Figures S4-S5**). The expansion and maintenance of mediated detoxification pathways in the spotted parrotfish appear to have worked in concert, resulting in higher enzyme levels for detoxification. Our results suggest that the spotted parrotfish is likely able to subsist on a highly specialized diet rich in allelochemicals and other toxic xenobiotic compounds. Whether these compounds are derived from cyanobacterial filaments or other ingested food items is unclear. Given the polyphyletic nature for many cyanobacterial genera [e.g. *Lyngbya* (Nicholson & Clements, 2023)], more studies are needed to corroborate these findings.

We find that ~25% (21 out of 83) of pigmentation-related gene (PRG) families were significantly expanded in the spotted parrotfish genome, the highest among all comparative teleost genomes in the present study (**Figure 3f**). We detected duplications of both *pax3a* and *pax3b* in the spotted parrotfish genome, resulting in four copies of *pax3* (versus two copies on average in other comparative teleosts; **Figure S6**), and a significant expansion of *chst8* in the spotted parrotfish (**Figure 3e**).

We identified 15 *scpp* genes in the spotted parrotfish. In contrast, between 12–20 *scpp* genes were identified for the other comparative teleost species **(Figure 4a)**. The difference in the number of *scpp* genes in these species is mostly attributed to the number of *scpp3* genes. For example, two *scpp3* genes were identified in the spotted parrotfish, while eight *scpp3* genes were found in the gilthead bream **(Figure 4a)**. Except for *scpp3* genes, the repertoire of *scpp* genes in the spotted parrotfish is similar to that of its phylogenetically closest relative, the humphead wrasse. We found no evidence of lineage-specific *scpp* gene duplications in the spotted parrotfish or any of the comparative labrid species, indicating that the suite of *scpp* genes is relatively well conserved for the family.

**3.3 | Genes that have undergone positive selection**

Of the 1,882 single-copy orthologs analysed, based on the branch-site mode, we inferred 143 (7.60%) genes that have undergone positive selection in the spotted parrotfish (**Table S1**). A GO enrichment analysis using clusterProfiler showed that the genes showing signals of positive selection in the parrotfish genome are significantly enriched (FDR-adjusted P-value < 0.2) and are involved in biological processes associated with cellular structure organization (e.g., actomyosin structure organization, GO:0031032; extracellular matrix organization, GO:0030198; respiratory tube development, GO:0030323 and locomotion, GO:0040011). Additionally, we also found that genes involved in focal adhesion, sex differentiation, and sex determination signalling pathways (KEGG: dre04510 *lamb2*, *lamc2*, *col9a2*, *kdr*, *pak7*, *igf1* and *rock1*) were significantly enriched (FDR-adjusted P-value < 0.2) based on the KEGG analysis.

**4 | DISCUSSION**

**4.1 | Ability to tolerate dietary toxins**

The diet of parrotfishes varies across species, but usually includes varying proportions of crustose coralline algae, dinoflagellates, and cyanobacterial filaments (Krayesky-Self et al., 2017; Clements & Choat, 2018; Fredericq et al., 2019; Nicholson & Clements, 2020; 2023). Atypical of most other parrotfishes, the diet of the spotted parrotfish also includes live coral, which, like its other dietary components, are also chemically defended (Gunthorpe & Cameron, 1990).

We identified expansions in several *cyp* and non-*cyp* mediated gene pathways that are known to play roles in detoxification and metabolism (Tseng et al., 2005; Boušová et al., 2015; Glisic et al., 2015; Liu et al., 2020; Contreras-Zentella et al., 2022; Liu et al., 2022). In most species of coral reef fishes, detoxification of allelochemicals is aided through enzymes encoded by genes in the cytochrome p450 (*cyp*) superfamily (Parkinson & Ogilvie, 2010; Maldonado et al., 2016). These genes have been identified from various organisms across the tree of life and encode heme-thiolate enzymes that play central roles in detoxification through oxidative metabolism of a range of compounds including xenobiotics (Kirischian et al., 2011). In addition, the tandem duplication and maintenance of 12 copies of *ces2b* in the spotted parrotfish genome could result in higher enzyme levels for detoxification. These genes encode for several carboxylesterases that are highly expressed in the liver and are responsible for the transformation and metabolism of xenobiotic compounds into more harmless metabolites (Tseng et al., 2005).

The gene expansions we have identified suggest that the spotted parrotfish is likely able to subsist on a highly specialized diet rich in allelochemicals and other toxic xenobiotic compounds. The species’ ability to not only metabolise these food items, but also mechanically extract them from within carbonite matrices, allows parrotfishes to occupy a trophic niche free from competition. However, whether these compounds are derived from cyanobacterial filaments or other ingested food items is unclear. Given the polyphyletic nature for many cyanobacterial genera [e.g. *Lyngbya* (Nicholson & Clements, 2023)], more studies are needed to corroborate these findings. Nonetheless, these results provide further insight into this extreme example of ecological niche partitioning as a result of dietary innovations within the Labridae.

**4.2 | Pigmentation, sex change, and their roles in sexual selection**

We identified expansions in a high proportion of PRG families (~25%) compared to the other genomes analysed, which are associated with pigmentation and sex change. In particular, *pax3*, for which occurs in four copies in the spotted parrotfish genome, is a member of the pairing box transcription factor family and is highly conserved during evolution (Barr et al., 1999). It is mainly expressed in muscle cells, skin, eyes, tail fin, and scales, and plays an important role in the differentiation, migration, and proliferation of melanin and xanthophore (Minchin & Hughes, 2008; Roberts et al., 2009). Further, *chst8* was significantly expanded in the spotted parrotfish, and is known to plays a role in the production of sex hormones in many organisms and is responsible for the sulfation and modification of luteinizing hormones (Dumitrescu & Collins, 2008; Cabral et al., 2012).

The expansion of gene families putatively involved in sex determination suggests an evolutionary role for sex determination and development of colour patterns in the spotted parrotfish. For example, in the widely sympatric Caribbean hamlet genus *Hypoplectrus*, selection acting on pigmentation (*sox10*), patterning (*hoxc13a*), photoreceptor development (*casz1*), and visual sensitivity (*sws* and *lws* opsins) were enough to facilitate speciation in syntopic species despite extremely low divergence elsewhere in the genome (Hench et al., 2019). Expansion of opsin *sws2b*, *lws1*, and *rh2* genes in at least one other labrid species (Liu et al., 2021) suggests that visual acuity and colour perception are key drivers of evolution in labrids, likely conferring an evolutionary advantage for these fishes with complex social biology that rely strongly on colouration patterns for correct identification of conspecifics (Kazancioglu et al., 2009).

We also detected positive selection in *pak7*, *igf1*, and *rock1*, genes that were putatively identified as being involved in sex differentiation or sex determination signalling pathways (Reinecke, 2010; Gegenhuber et al., 2022). Positive selection was also detected in genes putatively involved in gonadal development (*col9a2*) (Piprek et al., 2018), and social behaviours (*pak7*) (Strochlic et al., 2012). In other vertebrates, *igf1* has been shown to accelerate development of testicular tissue (Rahaie et al., 2018), whereas *rock1* has been shown to be tightly involved in oestrogen signalling (Huang et al., 2020). In labrids, sex change and sequential hermaphroditism is tightly regulated by social hierarchy, with sex change suppressed in the presence of a dominant terminal phased male. Because individuals frequently live in highly competitive and hierarchical social groups, it is critical that individuals rapidly assert and maintain behavioural dominance in order to achieve gonadal and morphological changes that lead to higher reproductive success (Lamm et al., 2015). In addition to morphological change in gonadal tissue, sex change is accompanied by development of secondary male sexual characteristics, such as development of male-specific colour patterns and external morphological traits. In some species, the initial dominance established by the highest-ranking female following removal of a terminal male to development of fully functional testes can happen in as little as 10 days (Warner & Swearer, 1991; Lamm et al., 2015). Together, these results provide support for sexual selection as being an important factor driving diversification in parrotfishes (Kazancioglu et al., 2009).

**4.3 | A putative role of poly-glutamine residues in teeth biomineralisation**

The beak-like teeth of parrotfishes consist of two upper dental plates, and two lower ones, each comprising about ~15 rows of teeth that are fused together. The biting end of these teeth are biomineralized with a fluorapatite enameloid with a remarkable hardness of up to 7.3 GigaPascals, making this one of the stiffest and hardest biominerals on earth7. Our analysis of genes, such as *scpp3,* could elucidate the evolution of this key trait.

Expression of *scpp3* genes is highly diverse across fishes. In pufferfishes for example, expression of *scpp3a* and *scpp3b* is concentrated in the oral epithelial cells covering the jaw and the pharyngeal epithelial cells overlying dental epithelium (Kawasaki et al., 2005). In the spotted gar, *scpp3cl* and *scpp3dl* are significantly expressed in the jaws but not in the skin, despite having ganoid scales covered with enamel (ganoin) (Mikami et al., 2022). Sticklebacks (family Gasterosteidae), while lacking scales entirely, possess six copies of *scpp3* genes **(Figure 4a)**. In contrast, *scpp3* is absent in seahorses, a group of fish lacking scales and teeth (Lin et al., 16; Zhang et al., 2020). These studies suggest that *scpp3* is involved in tooth formation, but not in scale formation, although the reason for variations in the number of *scpp3* genes remains to be determined.

Interestingly, while the repertoire of *scpp* genes and their copy numbers showed minimal variation across comparative labrid species, analysis of amino acid sequences encoded by *scpp* genes revealed large differences in one particular gene, *ambn*. In mammals, *ambn* encodes a dental enamel matrix protein. Similarly, *ambn* is expressed during dental and scale enamel formation in non-teleostean fishes such as gars and bichirs (Kawasaki et al., 2021; 2022). Although enamel was secondarily lost from teeth and scales in teleosts, *ambn* is still expressed in dental epithelial cells during teeth formation (Kawasaki et al., 2021), suggesting that *ambn* is somewhat involved in teeth formation of modern teleosts. The *ambn* gene in the spotted parrotfish encodes for 363 amino acids, while the *ambn* ortholog in its closest phylogenetic relative, the humphead wrasse **(Figure 4b)**, encodes 309 amino acids. This difference is largely due to the number of glutamine (Gln) residues encoded by exon 4 **(Figure 4c)**. In the spotted parrotfish, exon 4 of *ambn* encodes for 52 Gln residues, 12 of which are uninterrupted. In contrast, the *ambn* ortholog in the humphead wrasse encodes only 12 Gln residues. In other teleosts, exon 4 of *ambn* encodes 6–8 Gln residues in the gilthead seabream and stickleback. Curiously, the exon is missing entirely in pufferfish **(Figure 4a)**.

Comparison of exon 4 of *ambn* with comparative labrid species reveals that it encodes between 45–58 Gln residues in three Labrini species, namely the ballan wrasse, corkwing wrasse, and the cunner. In contrast, this exon encodes only 8–9 Gln residues in two other species, bluehead wrasse and sheepshead wrasse **(Figure 4c)**. Among non-labrid teleosts, exon 4 of *ambn* encodes 16 (zebrafish) or fewer Gln residues. As far as we are aware, a large exon 4 of *ambn* encoding for 26 or more Gln residues has not been reported previously. These findings are the first for the following labrids: spotted parrotfish (tribe Scarini: with 52 Gln residues); ballan wrasse, corkwing wrasse, and the cunner (*Tautogolabrus adspersus*) (tribe Labrini: with 45–58 Gln residues). Phylogenetic reconstruction suggests a sister relationship between the Scarini and the Cheilinini, with the two sharing a sister relationship to the Labrini(Hughes et al., 2022) **(Figure 4b)**. Given that poly-Gln residues were not detected in the humphead wrasse, it appears that a rare poly-Gln expansion encoded by exon 4 of *ambn* occurred once in the common ancestor of the Scarini + Cheilinini + Labrini, but was secondarily lost in the Cheilinini.

In vertebrates, mineralization and skeletal growth are accomplished through proteins containing polyproline repeat elements(Jin et al., 2009). In particular, glutamine plays an important role in biomineralization and in the formation of enamel and hydroxyapatite(Jin et al., 2009). Compaction of these polyproline motifs is especially important during enamel and apatite formation. Within these motifs, glutamine has been shown to greatly influence the compaction and function of these polyproline helices. Experimental replacement of even as few as 5 Gln residues with substitutes (e.g., alanine) has been shown to not only dramatically alter the diameter and density of these helices, but also reverse the effect of macromolecular compaction. These results indicate that Gln residues play a pivotal role in compaction of polyproline helices as they occur in many biological systems including biominerals(Jin et al., 2009).

We hypothesize that the poly-Gln expansion encoded by exon 4 of *ambn* in the parrotfish facilitates compaction of polyproline motifs expressed in the upper dentin layers of the parrotfish’s beak. However, why this poly-Gln expansion is retained in the Labrini is unclear. Members from this tribe typically feed on hard-shelled crustaceans and molluscs, and having highly biomineralized teeth may help facilitate breakdown of shelled prey(Ouannes-Ghorbel et al., 2006). However, the absence of poly-Gln expansion in other labrid groups that feed on similar prey items is curious (e.g., the sheepshead wrasse, tribe Hypsigenyini), suggesting multiple, independent evolutionary pathways for biomineralization of teeth within the Labridae.

**5 | CONCLUSIONS**

Our study provides a fine-grained approach in understanding the various evolutionary forces shaping the spotted parrotfish genome. Expansion and selection of detoxification gene families provide evidence for adaptive radiation as a result of ecological niche partitioning, allowing parrotfishes to occupy a highly unique trophic resources potentially rich in allelochemicals and xenobiotic compounds(Clements & Choat, 2018; Nicholson & Clements, 2020). Interestingly, metabolism of these toxic compounds was mediated by *cyp* and non-*cyp* pathways, with gene expansion, tandem duplication, and positive selection likely accounting for higher enzyme load in response to this dietary shift.

Typical of other labrids, we also find evidence of sexual selection and its role in the evolution of parrotfishes, particularly in the expansion of pigmentation-related gene families and genes encoding for sex hormones used in the biological process of sequential hermaphroditism. Together, these results suggest that while dietary shift is an important factor in the evolutionary history of the parrotfishes, there is likely a complex interplay between sexual selection and adaptive radiation(Kazancioglu et al., 2009). For example, we still do not fully understand the role of poly-Gln expansion in the *ambn* gene of the spotted parrotfish. While its role is putatively linked to teeth biomineralization, further studies are needed to ascertain this. Nonetheless, our assembled genome provides a first step in understanding the evolutionary complexity behind one of the most unique dietary specialists on coral reefs.

**Acknowledgements**

We thank A. Hay, S. Reader, and K. Parkinson for providing access to the tissue collections at the Australian Museum Research Institute. C.S. Foster and F. Jaya provided valuable feedback during the early stages of manuscript preparation. D. Earles and A. Hills provided live images. K. Evans provided micro-CT scans of *C. ocellatus*.

**Author’s contributions**

Y.K.T., K.K., S.Y.W.H., N.L. and S.H.F. devised the study. J.D.D. provided access to tissue samples. Y.K.T., Y.Z., K.M.E., G.C., and K.K. conducted the analyses. Y.K.T. wrote the manuscript with input from all authors. All authors approved the final version of the manuscript and agree to be held accountable for the content.

**Data availability statement**

All raw sequence data will be made available upon acceptance. Genome annotation files are deposited in Dryad (<https://doi.org/10.5061/dryad.j6q573nkz>) (for peer review: https://datadryad.org/stash/share/r2\_Vh7JTVN-tHTDCD\_ML-4BVUB95aT6jOaB-WojNiEg). The genome sequence of the spotted parrotfish is publicly available on GenBank [accession number XXX].

**Funding**

The study was funded by a University of Sydney and Fudan University partnership grant awarded to Y.K.T., S.Y.W.H., N.L. and S.H.F. S.H.F. is supported by grants from the National Key R&D Program of China (2020YFE0201600 and 2021YFC2500202), National Natural Science Foundation of China (31970563), the 111 Project (B13016), and Shanghai Municipal Science and Technology (2017SHZDZX01 and 19410741100). Y.K.T. is supported by the Chadwick Biodiversity Fellowship.

**References**

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology,* 215, 403–410.  
Amemiya, C.T. et al. (2013) The African coelacanth genome provides insights into tetrapod evolution. *Nature*, 496, 311–316.  
Aparicio, S. et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297, 1301–1310.  
Barr, F.G., Fitzgerald, J.C., Ginsberg, J.P., Vanella, M.L., Davis, R.J., & Bennicelli, J.L. (1999) Predominant expression of alternative Pax3 and Pax7 forms in myogenic and neural tumor cell lines. *Cancer Research,* 59, 5443–5448.  
Baxter, L.L., Watskins-Chow, D.E., Pavan, W.J., & Loftus, S.K. (2018) A curated gene list for expanding the horizons of pigmentation biology. *Pigment Cell Melanoma Research,* 32, 348–358.  
Bellwood, D.R., Goatley, C.H.R, Cowman, P.F., & Bellwood, O. (2015) The evolution of fishes on coral reefs: fossils, phylogenies, and functions. In: *Ecology of Fishes on Coral Reefs* (ed C. Mora). Pp 54–67. Cambridge, UK: Cambridge University Press.  
Bonaldo, R.M., Hoey, A.S., & Bellwood, D.R. (2014) The ecosystem roles of parrotfishes on tropical reefs. *Oceanography & Marine Biology: An Annual Review,* 52, 81–132.  
Boušová, I., Skálová, L., Souček, P., & Matouškovà, P. (2015) The modulation of carbonyl reductase 1 by polyphenols. *Drug Metabolism Rev*iews, 47, 520–533.  
Braasch, I. et al. (2016) The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nature*, 48, 427–437.  
Brawand, D. et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513, 375–381.  
Burbink, F. T., & Pyron, R.A. (2010) How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in New World ratsnakes (tribe Lampropeltini)? *Evolution*, 64, 934–943.  
Cabral, R.M., Kurban, M., Wajid, M., Shimomura, Y., Petukhova, L., & Christiano, A.M. (2012) Whole-exome sequencing in a single proband reveals a mutation in the chst8 gene in autosomal recessive peeling skin syndrome. *Genomics*, 99, 202–208.  
Cantalapiedra, C., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology & Evolution,* 38, 5825–5829.  
Capella-Gutiérrez, S., Silla-Martínez, J.M., & Gabaldón, T. (2009) trimAI: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973.  
Chen, Z., Pham, L., Wu, T.C., Mo, G., Xia, Y., Chang, P.L., Porter, D., Phan, T., Che, H., Tran, H., Bansal, V., Shaffer, J., Belda-Ferre, P., Humphrey, G., Knight, R., Pevzner, P., Pham, S., Wang, Y., & Lei, M. (2020) Ulutralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. *Genome Research*, 30, 898–909.   
Clements, K.D. & Choat, J.H. (2018) Nutritional ecology of parrotfishes (Scarinae, Labridae). In: *Biology of Parrotfishes* (eds A.S. Hoey & R.M. Bonaldo). Pp 42–68. Boca Raton, FL: CRC Press.

Contreras-Zentella, M.L., Villalobos-García, D., & Hernández-Muños, R. (2022) Ethanol metabolism in the liver, the induction of oxidant stress, and the antioxidant defence system. *Antioxidant*, 11, 1258.   
Criscuolo, A. & Garibaldo, S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology,* 10, 210.

Cunningham, F. et al. (2022) Ensembl 2022. *Nucleic Acids Research*, 50, 988–995.

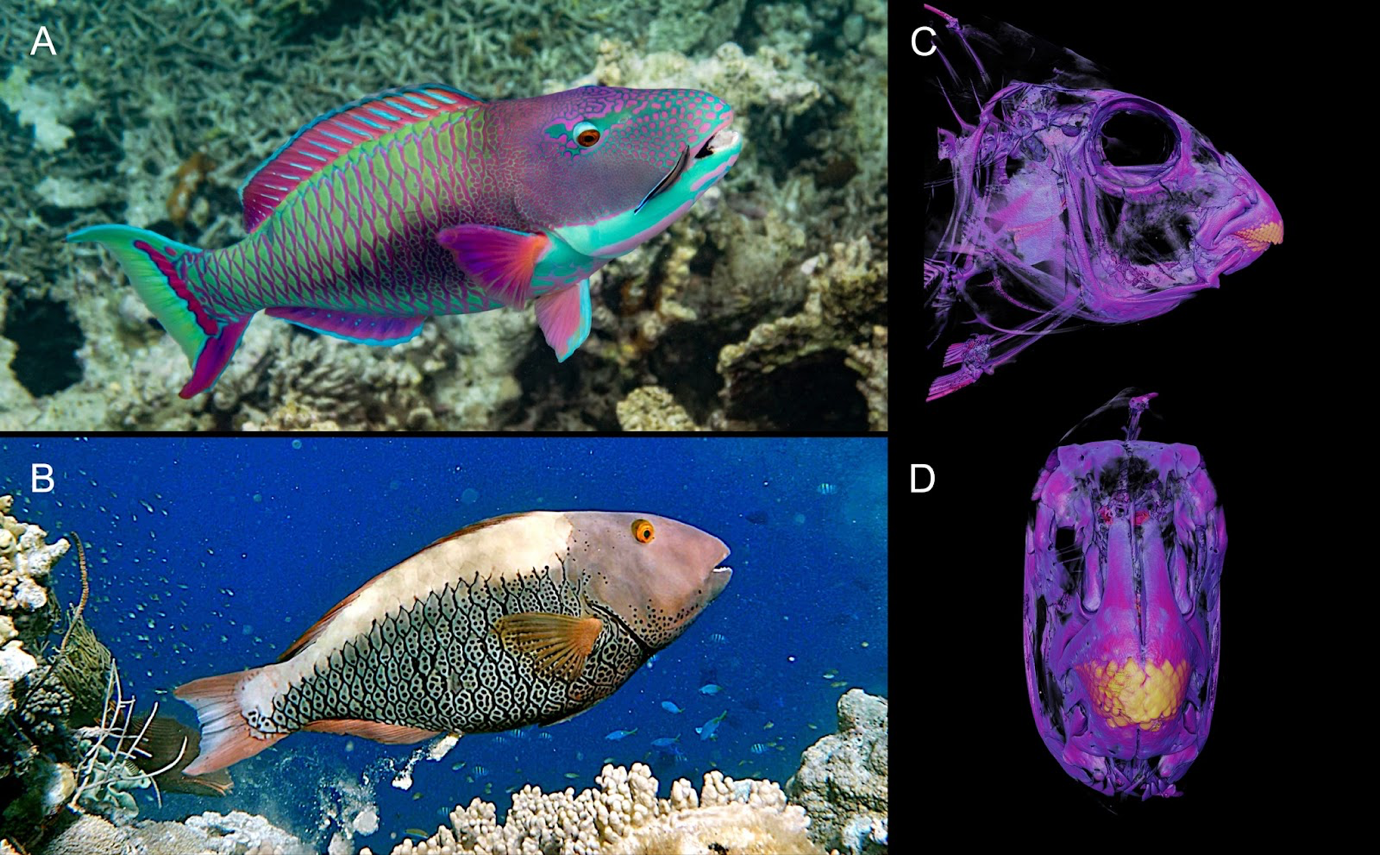
Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., & Li, H. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, 10, giab008.  
DeMaere, M.Z. & Darling, A.E. (2021) qc3C: Reference-free quality control for Hi-C sequencing data. *PLOS Computational Biology,* 17, e1008839.  
Dumitrescu, C.E. & Collins, M.T. (2008) McCune-Albright syndrome. *Orphanet Journal of Rare Diseases,*3, 12.  
Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., & Aiden, E.L. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems*, 3, 99–101.  
Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology,* 20, 238.  
Evans, K.M., Williams, K.L., & Westneat, M.W. (2019) Do coral reefs promote morphological diversification? Exploration of habitat effects on labrid pharyngeal jaw evolution in the era of big data. *Integrative & Comparative Biology*, 59, 696–704.  
Fan, G. et al. (2020) Initial data release and announcement of the 10,000 fish genomes project (Fish10K). *GigaScience*, 9, giaa080.  
Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., & Smit, A.F. (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Science, USA,* 117, 9451–9457.  
Fredericq, S., Krayesky-Self, S., Sauvage, T., Richards, J., Kittle, R., Arakaki, N., Hickerson, E., & Schmidt, W.E. (2019) The critical importance of rhodoliths in the life cycle completion of both macro- and microalgae, and as holobionts for the establishment and maintenance of marine biodiversity. *Frontiers in Marine Sci*ence, 5, 1–17.  
Garrison, E. & Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint: 1207.3907.  
Gegenhuber, B., Wu, M.V., Bronstein, R., & Tollkuhn, J. (2022) Gene regulation by gonadal hormone receptors underlies brain sex differences. *Nature*, 606, 153–159.  
Ghurye, J., Rhie, A., Walenz, B.P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A.M., & Koren, S. (2019) Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Computational Biology,* 15, e1007273.  
Glisic, B., Mihaljevic, I., Popovic, M., Zaja, R., Loncar, J., Fent, K., Kovacevic, R., & Smital, T. (2015) Characterisation of glutathione-S-transferases in zebrafish (*Danio rerio*). *Aquatic Toxicology,* 158, 50–62.  
Grant, P.R. (2013) Adaptive radiation. In: *The Princeton Guide to Evolution* (eds J.B. Losos, D.A. Baum, D.F. Futuyma, H.E. Hoekstra, R.E. Lenski, A.J. Moore, C.L. Peichel, D. Schluter, & M.C. Whitlock). pp. 559–566. New Jersey, USA: Princeton University Press.   
Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36, 2896–2898.  
Gunthorpe, L. & Cameron, A.M. (1990) Toxic exudate from the hard coral *Goniopora tenuidens*. *Toxicon*, 28, 1347–1350.  
Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29, 1072–1075.  
Hench, K., Vargas, M., Höppner, M.C., Mcmillan, W.O., & Puebla, O. (2019). Inter- chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evol*ution, **3,** 657–667.  
Howe, K. et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496, 498–503.  
Huang, Z.X., Mao, X.M., Wu, R.F., Huang, S.M., Ding, X.Y., Chen, Q.H., & Chen, Q.X. (2020) RhoA/ROCK pathway mediates the effect of oestrogen on regulating epithelial-mesenchymal transition and proliferation in endometriosis. *Journal of Cellular & Molecular Medicine,* 24, 10693–10704.  
Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., & Bork, P. 2019 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research,* 47, 309–314.  
Huertas, V., & Bellwood, D.R. (2020) Trophic separation in planktivorous reef fishes: a new role for mucus? *Oecologia*. 192, 813–822.  
Hughes, L., Nash, C.M., White, W.T., & Westneat, M.W. (2022) Concordance and discordance in the phylogenomics of the wrasses and parrotfishes (Teleostei: Labridae). *Systematic Biology,* Syac072.

Jin, T., Ito, Y., Luan, X., Dangaria, S., Walker, C., Allen, M., Kulkarni, A., Gibson, C., Braatz, R., Liao, X., et al. (2009) Elongated polyproline motifs facilitate enamel evolution through matrix subunit compaction. *PLOS Biology,* 7, e10000262.  
Jones, F.C. et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55–61.  
Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., & Petrov, A.I. (2018) Rfam 13.0: shifting to a genome- centric resource for non-coding RNA families. *Nucleic Acids Research,* 46, 335–342.  
Kasahara, M. et al. (2007) The medaka draft genome and insights into vertebrate genome evolution. *Nature*, 447, 714–719.  
Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology & Evolution,* 30, 772–780.   
Kawasaki, K., Keating, J.N., Nakatomi, M., Welten, M., Mikami, M., Sasagawa, I., Puttick, M.N., Donoghue, P.C.J., & Ishiyama, M. (2021) Coevolution of enamel, ganoin, enameloid, and their matrix SCPP genes in osteichthyans. *iScience*, 24, 102023.  
Kawasaki, K., Sasagawa, I., Mikami, M., Nakatomi, M., & Ishiyama, M. (2022) Ganoin and acrodin formation on scales and teeth in spotted gar: a. vital role of enamelin in the unique process of enamel mineralization. *Journal of Experimental Zoolology B: Molecular Development & Evolution*, 2022, 1–14.  
Kawasaki, K., Suzuki, T., & Weiss, K.M. (2005) Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proceedings of the National Academy of Science, USA*, 102, 18063–18068.   
Kazancioglu, E., Near, T.J., Hanel, R., & Wainwright, P.C. (2009) Influence of sexual selection and feeding functional morphology on diversification rate of parrotfishes (Scaridae). *Proceedings of the Royal Society B: Biology,* 276, 3439–3446.   
Kirischian, N., McArthur, A.G., Jesuthasan, C., Krattenmacher, B., & Wilson, J.Y. (2011) Phylogenetic and functional analysis of the vertebrate cytochrome P450 2 family. *Journal of Molecular Evolution,* 72, 56–71.  
Klaassen, C.D. & Watskins, J.B. (2010) Casarett and Doull’s Essentials of Toxicology. 2nd edition New York, USA: McGraw-Hill. 472 pp.  
Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research,* 27, 722–736.  
Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019) RAxML-NG: a gast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35, 4453–4455.  
Krayesky-Self, S., Schmidt, W.E., Phung, D., Henry, C., Sauvage, T., Camacho, O., Felgenhaeuer, B.E., & Fredericq, S. (2017) Eukaryotic life inhabits rhodolith-forming coralline algae (Hapalidiales, Rhodophyta), remarkable benthic microhabitats. *Scientific Reports,* 7, 45850.  
Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018) MEGA X: molecular evolutionary genetic analysis across computing platforms. *Molecular Biology & Evolution,* 35, 1547–1549.  
Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H-H., Rognes, T., & Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research,* 35, 3100–3108.  
Lamm, M.S., Liu, H., Gemmell. N.J., & Godwin, J. (2015) The need for speed: neuroendocrine regulation of socially-controlled sex change. *Integrative & Comparative Biology,* 55, 1–16.  
Lander, et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860– 921.  
Letunic, I., & Bork, P. (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research,* 49, 293–296.  
Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094–3100.  
Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.  
Lie, K.K., Tørrensen, O.K., Solbakken, M.H., Rønnestad, I., Tooming-Klunderud, A., Nederbragt, A.J., Jentoft, S., & Sæle, Ø. (2018) Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (*Labrus bergylta*) genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish. *BMC Genomics*, 19, 186.  
Lin, Q., Fan, S., Zhang, Y., Xu, M., Zhang, H., Yang, Y., Lee, A.P., Woltering, J.M., Ravi, J., & Gunter, H.M. et al. (2016). The seahorse genome and the evolution of its specialized morphology. *Nature*, **540**, 395–399.  
Liu, D., Wang, X., Guo, H., Zhang, X., Zhang, M., & Tang, W. (2021) Chromosome-level genome assembly of the endangered humphead wrasse *Cheilinus undulatus*: Insight into the expansion of opsin genes in fishes. *Molecular Ecology Resources*, 21, 2388– 2406.  
Liu, S., Wang, Z., Tian, X., & Cai, W. (2020) Predicting the effects of CYP2C19 and carboxylesterases on Vicagrel, a novel P2Y12 antagonist, by physiologically based pharmacokinetic/pharmacodynamic modelling approach. *Frontiers in Pharmacology,* 11, 591854.   
Lowe, T.M. & Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research,* 25, 955–964.  
Löytynoja, A. & Goldman, N. (2010) webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*, 11, 579.  
Marcus, M.A., Amini, S., Stifler, C.A., Sun, C-Y, Tamura, N., Bechtel, H.A., Parkinson, D.Y., Barnard, H.S., Zhang, X.X.X., Chua, J.QI., et al. (2017) Parrotfish teeth: stiff biominerals whose microstructure makes them tough and abrasion-resistant to bite stony corals. *ACS Nano*, 11, 11856–11865.   
Maldonado, A., Johnson, A., Gochfeld, D., Slattery, M., Ostrander, G.K., Bingham, J-P., & Schlenk, D. (2016) Hard coral (*Porites lobata*) extracts and homarine on cytochrome P450 expression in Hawaiian butterflyfishes with different feeding strategies. *Comparative Biochemistry & Physiology Part C: Toxicology & Pharmacology*, 179, 57–63.

Mattingsdal, M., Jentoft, S., Tørrensen, O.K., Knutsen, G., Hansen, M.M., Robalo, J.I., Zagrodzka, Z., André, C., & Gonzalez, E.B. (2018) A continuous genome assembly of the corkwing wrasse (*Symphodus melops*). *Genomics*, 110, 399–403.  
Mendes, F.K., Vanderpool, D., Fulton, B., & Hahn, M.W. (2020) CAFÉ 5 models variation in evolutionary rates among gene families. *Bioinformatics*, 36, 5516–5518.  
Mikami, M., Ineno, T., Thompson, A.W., Braasch, I., Ishiyama, M., & Kawasaki, K. (2022) Convergent losses of SCPP genes and ganoid scales among non-teleost actinopterygians. *Gene*, 811, 146091.  
Minchin, J.E. & Hughes, S.M. (2008) Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. *Developmental Biology,* 317, 508–522.

Nawrocki, E.P. & Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29, 2933–2935.  
Nei, M. & Kumar, S. (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York.   
Nelson, D.R. (2009) The cytochrome P450 homepage. *Human Genomics*, 4, 59–65.  
Nicholson GM, Clements KD. (2020) Resolving resource partitioning in parrotfishes (Scarini) using microhistology of feeding substrata. Coral Reefs, 39, 1313–1327.  
Nicholson, G.M. & Clements, K.D. (2023) Micro-photoautotroph predation as a driver for trophic niche specialization in 12 syntopic Indo-Pacific parrotfish species. *Biological Journal of the Linnean Society*, 139, 91–114.  
O’Leary, N.A. et al. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research,* 44, 733– 745.  
Ouannes-Ghorbel, A., & Bouain, A. (2006) The diet of the peacock wrasse, *Symphodus* (*Crenilabrus*) *tinca* (Labridae), in the southern coast of Tunisia. *Acta Adriatica,* 47, 175–182.  
Pan, H., Yu, H., Ravi, V., Li, C., Lee, A.P., Lian, M.M., Tay, B-H., Brenner, S., Wang, J., Yang, H., Zhang, G., & Venkatesh, B. (2016) The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *GigaScience*, 5, 36.  
Parkinson, A. & Ogilvie, B.W. (2010) Biotransformation of xenobiotics. In: Casarett and Doull’s Essentials of Toxicology. 2nd edition (eds C.D. Klaassen & J.B. Watkins). McGraw-Hill, New York. 472 pp.  
Pérez-Sánchez, J., Naya-Català, F., Soriano, B., Piazzon, M.C., Hafez, A., Gabaldón, T., Llorens, C., Sitjà-Bobadilla, A., & Calduch-Giner, J.A. (2019) Genome sequencing and transcriptome analysis reveal recent species-specific gene duplications in the plastic gilthead sea bream (*Sparus aurata*). *Frontiers in Marine Science,* 6, 760.  
Piprek, R.P., Kolasa, M., Podkowa, D., Kloc, M., & Kubiak, J.Z. (2018) Transcriptional profiling validates involvement of extracellular matrix and proteinases genes in mouse gonad development. *Mechanical Development,* 149, 9–19.  
Price, S.A., Holzman, R., Near, T.J., Wainwright, P.C. (2011) Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecology Letters,* 14, 462–469.  
Quinlan, A., & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.  
Rahaie, A., Toghyani, M., & Eghbalsaied, S. (2018) Cotreatment of igf1 and fadrozole upregulates the expression of rspo1, sox9, and amh in chicken embryos. *Cells Tissues Organs*, 206, 218–228.  
Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E.J.P. (2011) MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLOS ONE*, 6, e22594.  
Reinecke, M. (2010) Influences of the environment on the endocrine and paracrine fish growth hormone – insulin-like growth factor-I system. *Journal of Fish Biology,* 76, 1233–1254.  
Roberts, R.B., Ser, J.R., & Kocher, T.D. (2009) Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science*, 326, 998–1001.  
Schluter, D. (2000) Ecological character displacement in adaptive radiation. *The* *American Naturalist,* 156, 4–16.  
Salamov, A.A. & Solovyev, V.V. (2000) Ab initio gene finding in *Drosophila* genomic DNA. *Genome Research*, 10, 516–522.  
Seppey, M., Manni, M., & Zdobnov, E.V. (2019) BUSCO: Assessing genome assembly and annotation completeness. In: *Gene Prediction*. Methods in Molecular Biology (ed M. Kollmar). Pp. 227–245. New York, USA: Humana.

Skinner, M.P., Brewer, T.D., Johnstone, R., Fleming, L.E., & Lewis, R.J. (2011) Ciguatera fish poisoning in the Pacific Islands (1998–2008). *PLOS Neglected Tropical Diseases,* 12, e1416.  
Smit, A., Hubley, R., & Green, P. (2015) RepeatMasker Open-4.0. Retrieved from <http://www.repeatmasker.org> (17 June 2022, date last accessed).  
Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology,* 7, S10.  
Streeman, J.T., Alfaro, M.E., Westneat, M.W., Bellwood, D.R., & Karl, S.A. (2002) Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. *Evolution*, 56, 961–971.  
Strochlic, T.I., Concilio, S., Viaud, J., Eberwine, R.A., Wong, L.E., Minden, A., Turk, B.E., Plomann, M., Peterson, J.R. (2012) Identification of neuronal substrates implicates pak5 in synaptic vesible trafficking. *Proceedings of the National Academy of Science, USA*, 109, 4116–4121.  
Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M., & Paterson, A.H. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research,* 18, 1944–1954.  
Tong, C., Li, M., Tang Y., & Zhao K. (2021) Genomic signature of shifts in selection and alkaline adaptation in highland fish. *Genome Biology and Evolution*, 13, evab086.  
Tseng, H.P., Hseu, T.H., Buhler, D.R., Wang, W.D., & Hu, C.H. (2005) Constitutive and xenobiotics-induced expression of a novel CYP3A gene from zebrafish larva. *Toxicology & Applied Pharmacology,* 205, 247–258.   
Warner, R.R., &Robertson, D.R. (1978) Sexual patterns in the labroid fishes of the western Caribbean, I: the wrasses (Labridae). *Smithsonian Contributions to Zoology,* 254, 1– 22.  
Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J.M., & Birol, I. (2015) LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience*, 4, 35.  
Westneat, M.W., Alfaro, M.E., Wainwright, P.C., Bellwood, D.R., Grubich, J.R., Fessler, J.L., Clements, K.D., & Smith, L.L. (2005) Local phylogenetic divergence and global evolutionary convergence of skull function in reef fishes of the family Labridae. *Proceedings of the Royal Society B: Biology,* 272, 993–1000.  
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhang, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, 2, 100141.  
Yang, Z. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology & Evolution,* 15, 568–573.  
Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology & Evolution,* 24, 1586–1591.  
Yang, Z. & Nielsen, R. (19980 Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution,* 46, 409–418.  
Yeo, S., Coombe, L., Warren, R.L., Chu, J., & Birol, I. (2018) ARCS: scaffolding genome drafts with linked reads. *Bioinformatics*, 34, 725–731.  
Zhang, J., Qi, J., Shi, F., Pan, H., Liu, M., Tian, R., Geng, Y., Li, H., Qu, Y., Chen, J., Seim, I., & Li, M. (2020) Insights into the evolution of neoteny from the genome of the Asian icefish *Protosalanx chinensis*. *iScience*, **23**, 101267.  
Zhang, J., Nielsen, R., & Yang, Z. (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology & Evolution,* 22, 2472–2479.

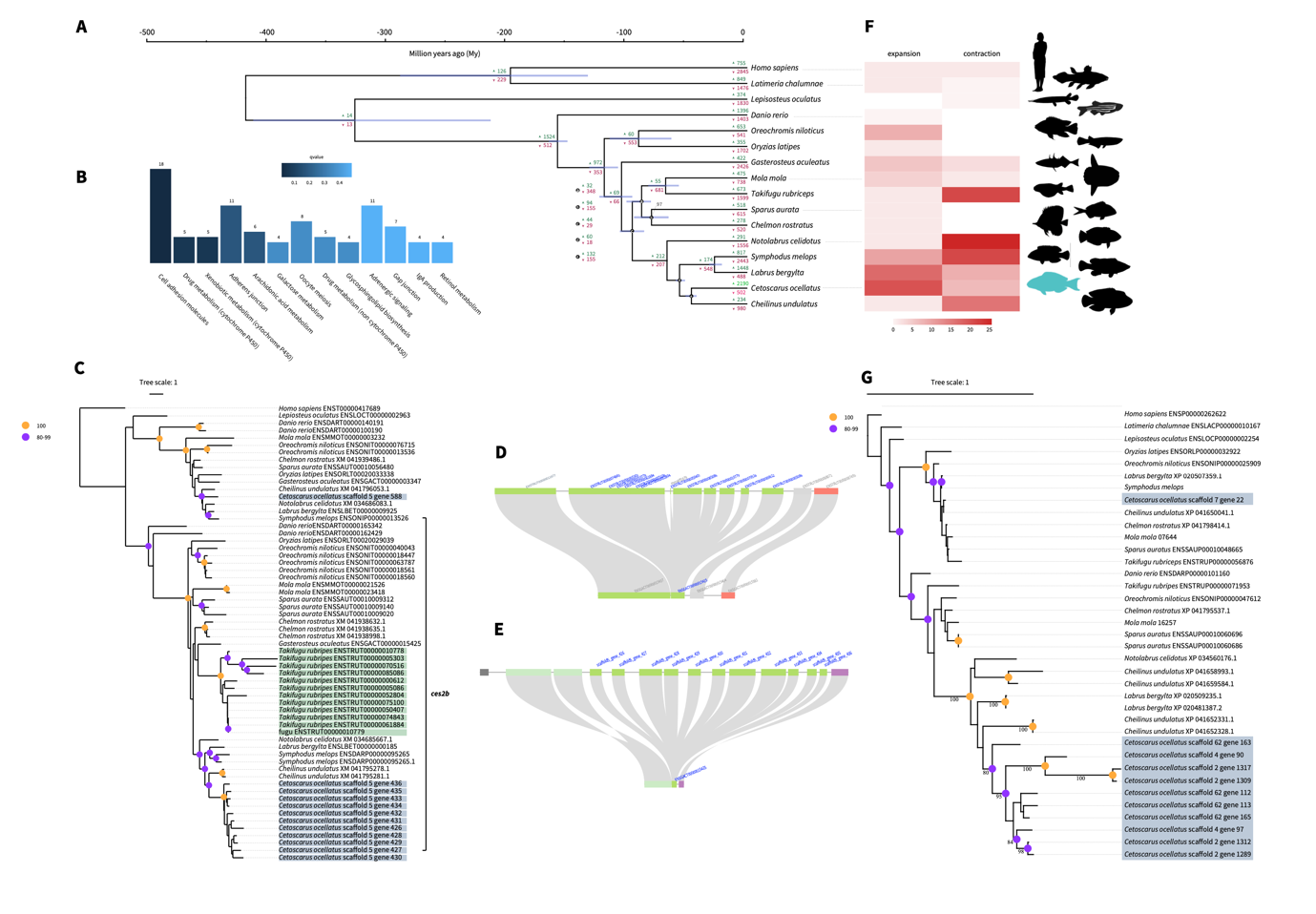


**FIGURE 1** *Cetoscarus ocellatus* (spotted parrotfish) demonstrating pronounced dimorphism in **(a)** male and **(b)** female body coloration. The beak-like dentition typical of parrotfishes is highlighted in the micro-CT scans **(c)** and **(d)**. Specimen registration number FMNH 110797. Photographs by David Earles **(a)**, Adelma Hills **(b)**, and Kory Evans (c) and (d).

A picture containing diagram, screenshot, line, plot

Description automatically generated

**FIGURE 2** Transposable element (TE) evolution across four labrid genomes. **(a)**: TE composition in the spotted parrotfish. TE expansion histories in the: **(b)** spotted parrotfish; **(c)** humphead Maori wrasse; **(d)** corkwing wrasse; and **(e)** ballan wrasse. X and Y axes indicate CpG-adjusted Kimura substitution levels and percentage of TEs, respectively. Red arrows indicate independent waves of TE expansion in the spotted parrotfish and humphead Maori wrasse genomes respectively.

**  
FIGURE 3** **(a)**: Time-calibrated phylogeny and gene families for various chordates. The coelacanth (Sarcopterygii), gar (non-teleostean Actinopterygii), and human (Mammalia) were used as outgroups. Unless specifically stated, all nodes had support values of 100%. Gene family expansion and contraction for each species and the MRCA are represented by arrows and values in green (for expansion) and red (for contraction), respectively. **(b)**: Enrichment analysis of KEGG pathway for 572 expanded gene families in the spotted parrotfish. X axis indicates the pathway name, Y axis indicates the number of gene families enriched in each pathways. The colour bar indicates the q-value. **(c)**: Maximum-likelihood phylogeny of *ces2* gene family in sixteen species of comparative chordates. **(d)**: Local synteny plot of *ces2b* from fugu (above) and stickleback (below). **(e**): Local synteny plot of *ces2b* from spotted parrotfish (above) and stickleback (below). **(f)**: Heatmap of expanded or contracted gene families related to pigmentation genes, the numbers indicate the expanded or contracted gene family numbers in the respective species of the phylogeny. The coloured bar indicates the range of the gene family number. **(g)**: Maximum-likelihood phylogeny of *chst8* gene family. The orange circle and purple circle indicate bootstrap support 100 and bootstrap support from 80 to 99 respectively. Bootstrap values (1000 bootstrap replicates) are reported as percentages.

A picture containing text, screenshot, diagram, parallel

Description automatically generated

**FIGURE 4** **(a):** Genomic arrangement of *scpp* genes in seven teleost fishes. Frags in different colors represent the order and transcriptional directions of *scpp* genes encoding proline/glutamine (P/Q)-rich proteins (red), sparcl1 (yellow), acidic proteins (blue) (Mikami et al., 2022), and other genes (black), as summarized in the legend. Numbered horizontal lines represent chromosomes. In zebrafish (*Danio rerio*), the original *scpp* gene cluster on chromosome 1 and chromosome 23 was split by an inversion (arrow) and separated by a 6-megabase (Mb) sequence and a 1-Mb sequence, respectively. In the zebrafish genome, *idua* and *rchy1* are separated by an 80-kilobase (kb) sequence. **(b)**: Maximum-likelihood gene tree of *ambn* and **(c)** the total number of glutamine residues (Q) encoded by exon 4 for comparative labrid species. Support values at nodes correspond to bootstrap support. Amino acid sequences are schematically shown here (see Supplemental Information for full amino acid sequences) with Gln residues in green and non-Gln amino acids in blue.