GENETICS

Seadragon genome analysis provides insights into its phenotype and sex determination locus

Meng Qu[1,2†], Yali Liu[1,2†], Yanhong Zhang[1,2†], Shiming Wan[1,2†], Vydianathan Ravi[3†], Geng Qin[1,2†], Han Jiang[1,4], Xin Wang[1,2], Huixian Zhang[1,2], Bo Zhang[1,2], Zexia Gao[5], Ann Huysseune[6], Zhixin Zhang[7], Hao Zhang[1,2], Zelin Chen[1,2], Haiyan Yu[8], Yongli Wu[1,4], Lu Tang[1,4], Chunyan Li[1,2], Jia Zhong[1,2], Liming Ma[8], Fengling Wang[8], Hongkun Zheng[8], Jianping Yin[1], Paul Eckhard Witten[6], Axel Meyer[9]*, Byrappa Venkatesh[3]*, Qiang Lin[1,2,4]*

The iconic phenotype of seadragons includes leaf-like appendages, a toothless tubular mouth, and male pregnancy involving incubation of fertilized eggs on an open "brood patch." We de novo–sequenced male and female genomes of the common seadragon (*Phyllopteryx taeniolatus*) and its closely related species, the alligator pipefish (*Syngnathoides biaculeatus*). Transcription profiles from an evolutionary novelty, the leaf-like appendages, show that a set of genes typically involved in fin development have been co-opted as well as an enrichment of transcripts for potential tissue repair and immune defense genes. The zebrafish mutants for *scpp5*, which is lost in all syngnathids, were found to lack or have deformed pharyngeal teeth, supporting the hypothesis that the loss of *scpp5* has contributed to the loss of teeth in syngnathids. A putative sex–determining locus encoding a male-specific *amhr2y* gene shared by common seadragon and alligator pipefish was identified.

INTRODUCTION

Seadragons are members of the family Syngnathidae that also includes the more derived seahorses and pipefishes. They are fascinating animals because of their dragon-shaped body with special leaf-like appendages and spectacular coloration. They are considered to be "masters of camouflage" as they can deceive predators by changing their color pattern to resemble and blend with kelp and other seaweed. Like other syngnathids, seadragons have specialized morphological traits such as a tube-like toothless mouth, the absence of pelvic fins and scales, a bony armor covering the body, and "male pregnancy." Males of all syngnathids incubate the fertilized eggs on a specialized incubation area, and in the modern lineages, even in enclosed brood pouches, making them prime examples for sex-role reversal (*1*–*3*). On the basis of the position of the brood pouch, syngnathids have been divided into tail-brooders (subfamily Syngnathinae represented by seahorses, some pipefishes and seadragons) and trunk-brooders (subfamily Nerophinae represented by Manado pipefish) (*4*, *5*). Among the Syngnathinae, seahorses and pipefishes have fully or partially enclosed brood pouch, whereas the basal seadragons still incubate fertilized sticky eggs that are attached to an open brood patch under the males' tail (Fig. 1A) (*6*). The incubation of fertilized eggs on the open brood patch in seadragons is the ancestral form of male pregnancy (*4*, *7*). In teleost fishes, sex determination is highly variable, and several types of sex chromosomes and master sex–determining genes have been reported even among closely related species (*8*). However, the sex determination mechanisms in seadragons and other syngnathids are still not well understood. Seadragons are gonochoristic species with the sex ratio close to 1:1, and there is no evidence to show that sex can be influenced by any environmental factors (*9*).

To understand the molecular basis of the evolutionary novelties of seadragons and to identify their sex-determining locus, we generated high-quality, chromosome-level genome assemblies of a male and a female individual of the common seadragon (*Phyllopteryx taeniolatus*, also known as weedy seadragon) (Fig. 1A). In addition, we generated a chromosome-scale, genome assembly for the alligator pipefish (*Syngnathoides biaculeatus*), which, although a trunk-brooder (Fig. 1B), has intriguingly been phylogenetically grouped with the tail-brooding subfamily Syngnathinae (*5*, *10*).

RESULTS AND DISCUSSION

Genome features

PacBio technology was used to generate genome sequences for two common seadragons and a male alligator pipefish, and Nanopore was used for sequencing the female alligator pipefish genome. In addition, low-coverage (30×) genome sequences of two other male common seadragons were generated with Illumina technology (HiSeq 2500) (tables S1 to S4). Hi-C reads were used to scaffold the contig-level assemblies resulting in chromosome-level genome assemblies comprising 23 chromosome-level scaffolds for each of these fishes (figs. S2 and S3). We made pairwise comparisons of the syntenic relationship between common seadragon, alligator pipefish, and lined seahorse (2n = 22) and found prevalent rearrangements in all three pairs of comparison (fig. S4). The assembled genome size of the male and female common seadragon is ~659 and ~663 Mb, respectively, with high contig N50 values (10.0 and 12.1 Mb), whereas the assembled genome size of the male and female alligator pipefish is ~637 and ~648 Mb, respectively, with even higher contig N50 values

[1]CAS Key Laboratory of Tropical Marine Bio-Resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, 510301 Guangzhou, China. [2]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), 511458 Guangzhou, China. [3]Institute of Molecular and Cell Biology, A*STAR, 138673 Biopolis, Singapore. [4]University of Chinese Academy of Sciences, 100101 Beijing, China. [5]College of Fisheries, Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, Huazhong Agricultural University, 430070 Wuhan, China. [6]Department of Biology, Ghent University, Ledeganckstraat 35, B-9000 Ghent, Belgium. [7]Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Minato, Tokyo, Japan. [8]Biomarker Technologies Corporation, 101300 Beijing, China. [9]Department of Biology, University of Konstanz, 78464 Konstanz, Germany.
*Corresponding author. Email: linqiang@scsio.ac.cn (Q.L.); mcbbv@imcb.a-star.edu.sg (B.V.); axel.meyer@uni-konstanz.de (A.M.)
†These authors contributed equally to this work.

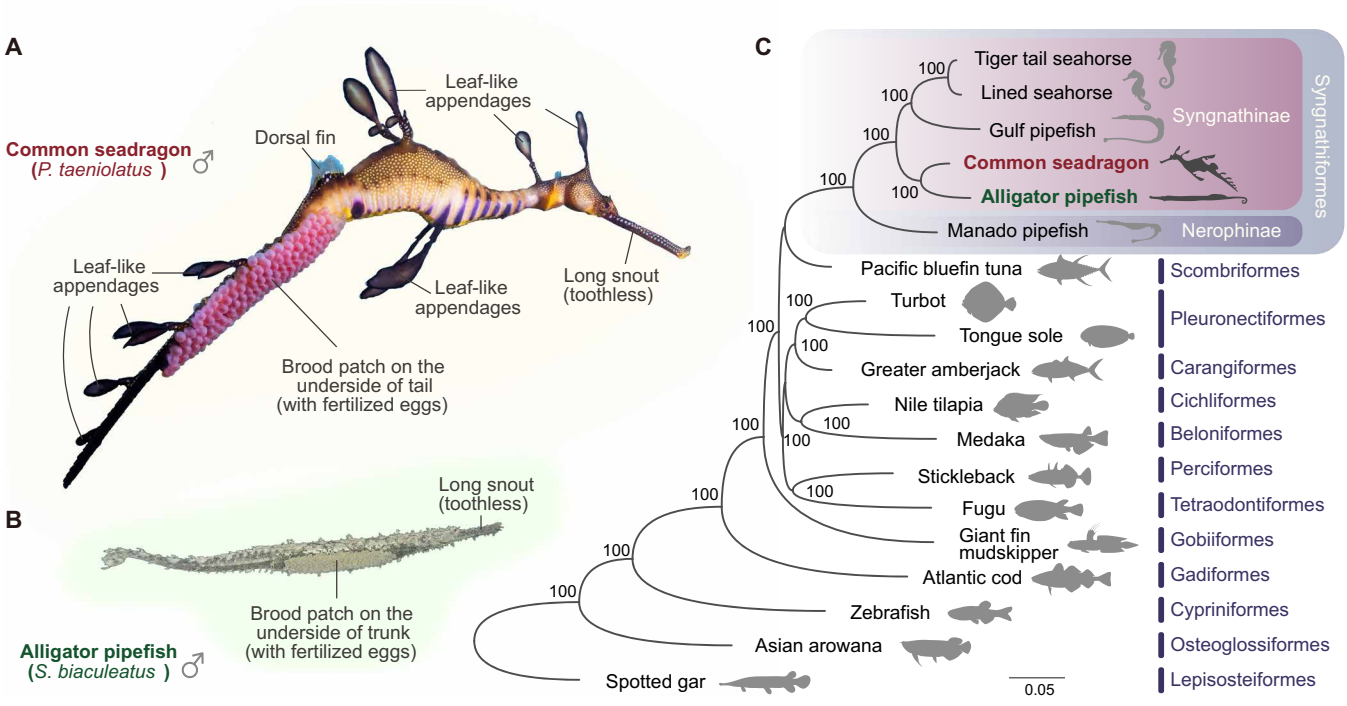

Fig. 1. Key features of the common seadragon (*P. taeniolatus*) and the alligator pipefish (*S. biaculeatus*) and their phylogenetic positions. (**A**) The common seadragon has a dragon-shaped body with many special leaf-like appendages and long tubular snout lacking teeth. Common seadragon males incubate eggs on a brood patch on the underside of the tail. (**B**) A male alligator pipefish showing eggs incubated on a brood patch on the underside of trunk. (**C**) Phylogenetic tree of 19 ray-finned fishes. Bootstrap values below 80 are not shown.

(18.0 and 21.0 Mb) (table S7). To evaluate the quality of these genome assemblies, we performed BUSCO analyses and obtained high scores (94.00 to 94.40%) (table S8). Analysis of gene family evolution identified 31 and 33 expanded gene families in the common seadragon and alligator pipefish, respectively (fig. S5).

The genome of the common seadragon is larger than most sequenced syngnathids, containing a high level of transposable elements (TEs) (~56%) and other interspersed repeats (table S11), with the Tc1/mariner transposon being the most dominant element (~55.3% of TEs). The Tc1/mariner transposon was found to have undergone two rounds of notable expansion in the common seadragon after its split from the alligator pipefish. Tc1/mariner was also found to have undergone one expansion event in alligator pipefish and Manado pipefish, but not sharing with the common seadragon, nor with each other, while no distinct expansions were found in the two seahorse species (fig. S6).

Phylogenomics

Previous studies based on some molecular and morphological characters had proposed the division of syngnathid fishes into two subfamilies based on the placement of the brooding area—Syngnathinae and Nerophinae—comprising tail-brooders and trunk-brooders, respectively (*4*, *7*, *11*). However, the trunk-brooding alligator pipefish was inconsistently placed within the tail-brooder clade (*5*, *10*). To determine the phylogenetic position of the common seadragon and to confirm the apparent inconsistency of the alligator pipefish grouping within the tail-brooders, we performed phylogenomic analysis using whole-genome protein datasets from syngnathids and other ray-finned fish species. Our phylogenomic analysis also places the trunk-brooding alligator pipefish within the tail-brooding Syngnathinae clade, as a sister taxon to the common seadragon with maximal bootstrap support (Fig. 1C and fig. S7). The clade comprising common seadragon and alligator pipefish appears as a sister to the remaining Syngnathinae members analyzed in this study (Fig. 1C and fig. S7). Thus, the position of the trunk-brooding alligator pipefish within the tail-brooders appears to be correct and the basal placement of these two genera to the more derived syngnathids such as seahorses and pipefish permits to study evolutionary trends of many unique phenotypic and life history traits within these fascinating fishes.

The species divergence times estimated on the basis of eight calibration points showed that common seadragon and alligator pipefish diverged around 27.3 Ma ago [95% highest posterior density (HPD) 13.7 to 25.9 Ma ago]. The timeline for other syngnathids diversification are broadly consistent with previous studies (*12*–*15*): Syngnathiformes and Scombriformes diverged around 83.8 Ma ago (95% HPD 50.1 to 104.1 Ma ago), the two syngnathid subfamilies Syngnathinae and Nerophinae diverged around 49.0 Ma ago (95% HPD 48.0 to 50.0 Ma ago), and the genus *Syngnathus* and *Hippocampus* diverged around 37.3 Ma ago (95% HPD 34.6 to 40.7 Ma ago) (fig. S7).

Genes related to leaf-like appendages, an evolutionary key innovation

The common seadragon camouflage is remarkable as it blends astonishingly well with its surroundings because of its leaf-like appendages on its head, neck, abdomen, dorsal, and tail region. These structures are an adaptive evolutionary key innovation that helps the fish to mimic seaweeds (*16*). Although the leaf-like appendages look like fins, histological analysis showed that the long leaf-like structures lack fin rays and are mainly composed of a bone matrix,

along with connective tissues enriched in collagenous fibers (Fig. 2A). This may explain how the leaf-like appendages remain flexible and sway with the flow of seawater. Transcriptomic profiles identified 443 highly expressed genes in the leaf-like appendages [fragments per kilobase of exon per million fragments (FPKM) > 100] (table S17), mainly enriched in glycolysis and glucose metabolism processes. Of these, the 20 most highly expressed genes include not only *collagens*, *osteocalcin*, and *keratins*, which likely provide strength and elasticity to the leaf-like appendages (*17*, *18*), but also genes related to inflammatory activity, such as *lectins* (I, II, and III types), β*2-macroglobulin*, and *apolipoprotein E*, suggesting a role in immune defense for the leaf-like appendages (Fig. 2A and table S18) (*19*). The evolutionary origin of new organs often depends on the recruitment and co-option of genes that were originally expressed in other tissues (*20*). Comparative transcriptome analysis from leaf-like appendages (paired/unpaired), fin, skin, intestine, gill, muscle, liver, brain, and eye suggested that the maintenance of leaf-like appendages is largely supported by genes recruited from the development and maintenance of fins (Fig. 2B, fig. S11, and table S19). For example, genes such as *msx*, *dlx*, and *fgf* that play a role in fin development were co-opted to be expressed in leaf-like appendages (*21*, *22*). The co-option of existing genes for the evolution of a novel phenotype seems to be a common theme in evolution. The transcriptome analysis also revealed some differences between the leaf-like appendages and fins such as the lack of limb-specific *hox* genes (*hoxa13b* and *hoxd12a*) expression in leaf-like appendages (*23*, *24*). A suite of genes related to immunity, tissue regeneration, and repair (*ust2b*, *D2*, *umodl1*, *pon2*, and *nr0b2*) were found to be highly expressed in the accident-prone leaf-like appendages (fig. S12 and table S20). Further analysis identified 2375 differentially expressed genes (DEGs) between the leaf-like appendages and fins. These included extracellular matrix receptor interaction and focal adhesion pathway genes such as *collagens*, *laminins*, and *integrins* that highly expressed in leaf-like appendages (Fig. 2A) (*25*). In addition, several inflammation and injury repair–related genes including *tnfrsfs*, *chemokines*, *chemokine receptors*, and *protocadherins* also showed higher expression levels in leaf-like appendages, while the expression of *wnts* and *hh* signaling related genes were higher in fins (figs. S13 and S14 and table S21). It is possible that these genes might have to do with the ability to quickly heal and regenerate the leaf like-appendages that are often damaged during the attacks by predators.

Expansion of the *MHC I* genes

A marked expansion of *MHC I* genes (32 genes) was found in the common seadragon compared to the alligator pipefish (11 genes) and other syngnathids (7 to 14 genes) (fig. S15A). Moreover, 19 *MHC I* genes in the common seadragon genome were found to be tandemly organized into a cluster, and the RTE and Gypsy TEs are quite abundant in this region, although Tc1-Mariner is the dominating type of TE in the whole genome (fig. S15, B and C). RTE and Gypsy are both superfamilies of long terminal repeat (LTR) retrotransposons, which have been suspected to affect the expression pattern of closest genes (*26*). RNA sequencing (RNA-seq) analysis of the seadragon showed that most of *MHC I* genes were expressed at considerably higher levels in the leaf-like appendages, skin, and gills that are in contact with water, compared to the liver (fig. S16B). The expanded family of *MHC I* genes in common seadragon might provide additional immune protection to this fish that depends largely on its camouflage ability to evade predators and whose appendages are particularly prone to damage.

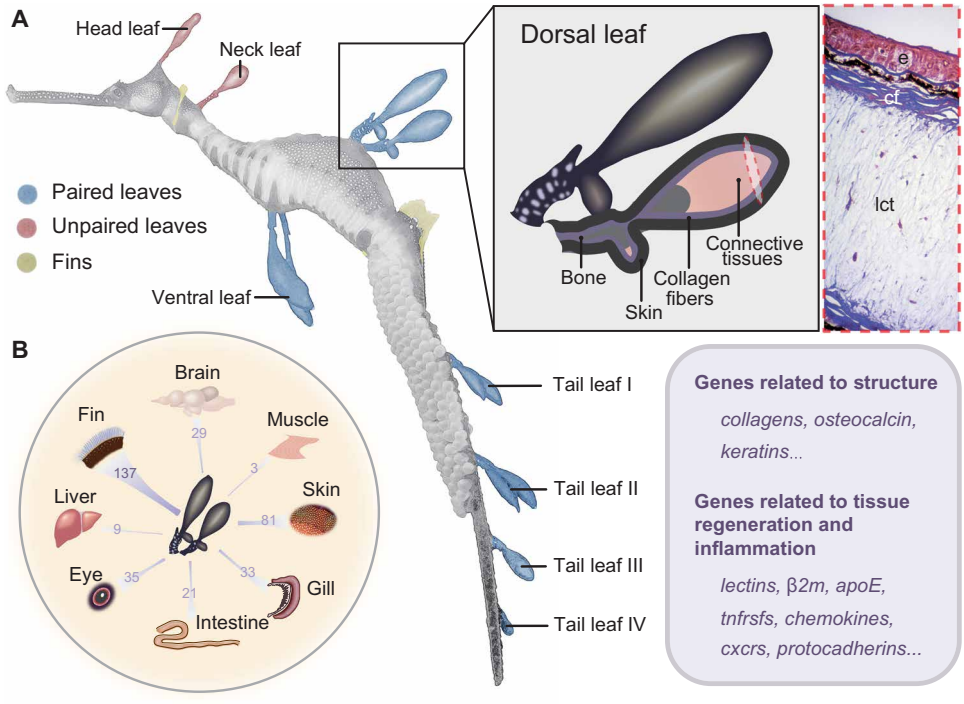

Fig. 2. Morphology and genetic makeup of the leaf-like appendages in the common seadragon (*P. taeniolatus*). (**A**) Location and structure of leaf-like appendages. The inset on the right shows a zoomed in schematic of the leaf-like appendages along with its transverse section. Genes expressed in the leaf-like appendages that are related to structure, tissue regeneration, and inflammatory response have been listed in the box below. e, epidermis; cf., collagen fiber; lct, loose connective tissues. (**B**) The number of genes recruited from other tissues for the maintenance of leaf-like appendages.

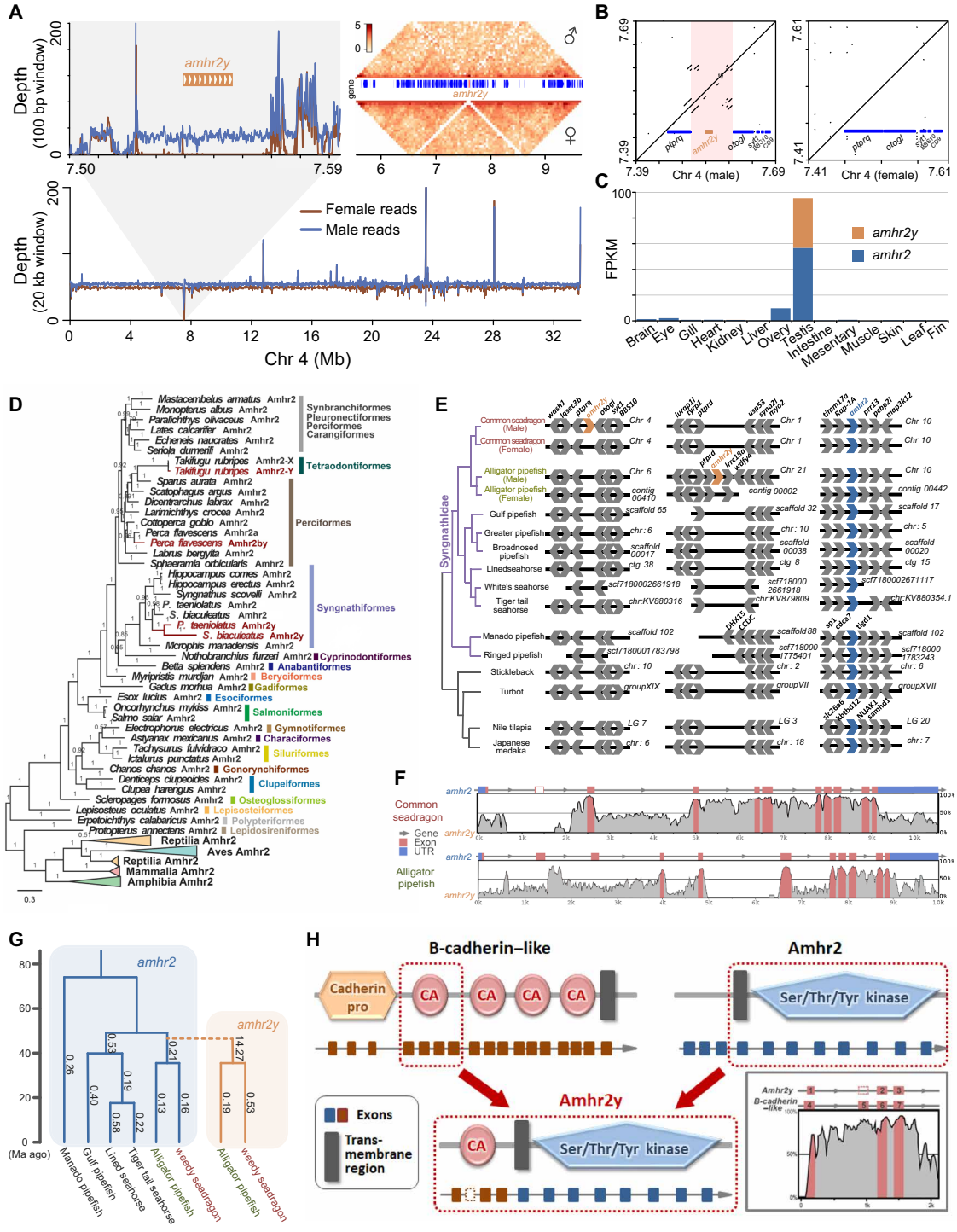

Fig. 3. Putative sex determination gene in the common seadragon (*P. taeniolatus*) and the alligator pipefish (*S. biaculeatus*). (**A**) Characterization of a Y-specific region in the common seadragon. Illumina data illustrating the relative read depth across a region of Chr4, in males (blue line) and females (red line) showing a coverage difference between the two; the Hi-C data heatmap of the Y-specific region showing the male-specific region, with half of the average read-coverage; gene annotation indicates only one gene (*amhr2y*, represented in orange). (**B**) Repetitive sequences present in the Y-specific region of the male genome but not in the female. (**C**) Tissue-specific expression of *amhr2y* and *amhr2* in the common seadragon. (**D**) Phylogenetic tree of amhr2 amino acid sequences in vertebrates. The Y-specific amhr2 (red) originated independently in pufferfish (*T. rubripes*), yellow perch (*P. flavescens*), and the common ancestor of common seadragon (*P. taeniolatus*) and alligator pipefish (*S. biaculeatus*). (**E**) Synteny analysis of *amhr2y* and *amhr2* loci in the common seadragon, alligator pipefish, and other teleosts. (**F**) Identity plot of the alignment of *amhr2y* gene with the autosomal *amhr2* gene sequence in the common seadragon and alligator pipefish. (**G**) The duplication event of *amhr2y* is shown as a dashed line in the phylogenetic tree. Clades of *amhr2* and *amhr2y* are in blue and orange, respectively. The ω values are shown at the right of each branch. (**H**) Schematic depiction of the events leading to the formation of the alligator pipefish *amhr2y* gene, which is likely to be the result of fusion between *B-cadherin–like* and *amhr2* genes.

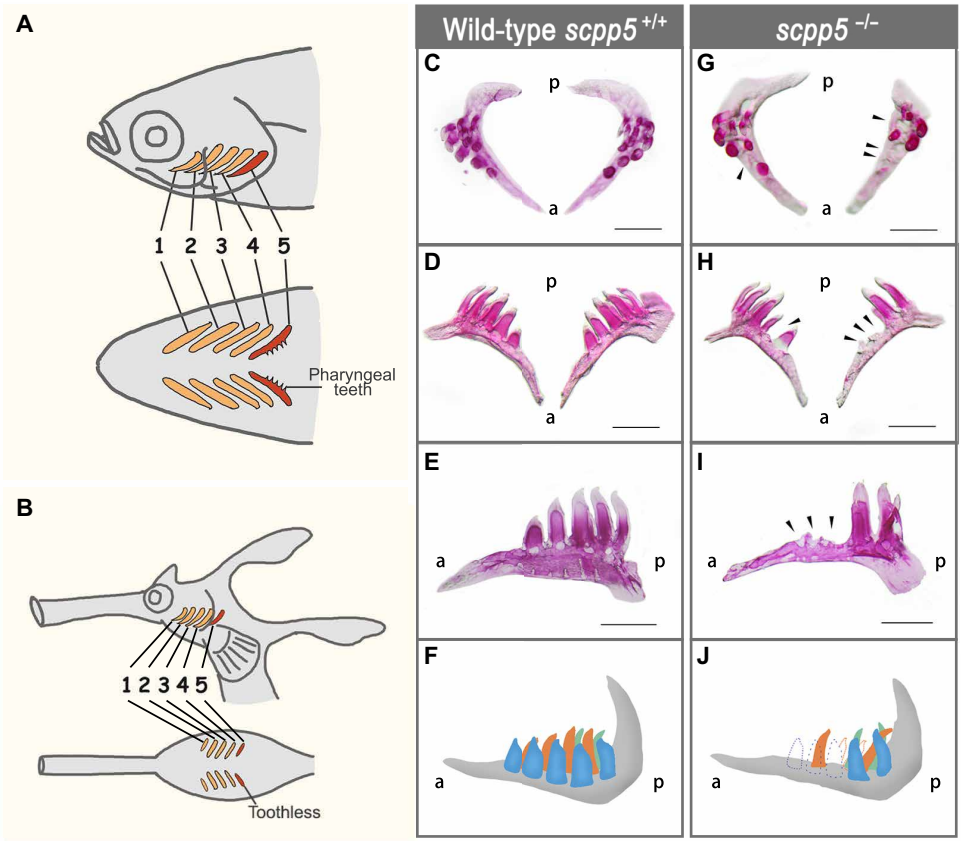

Fig. 4. Pharyngeal tooth phenotypes in zebrafish *scpp5* homozygous mutants. (**A**) Schematic diagram showing the teeth located on the fifth gill arch in zebrafish. (**B**) Schematic diagram showing five pairs of toothless pharyngeal arches in the common seadragon. (**C** to **J**) Comparison of number and distribution of pharyngeal teeth in mutant (*scpp5*$^{-/-}$; G, H, and I) versus wild-type (*scpp5*$^{+/+}$) fish (C, D, and E). Eleven of 24 (45.8%) F3 homozygous *scpp5*$^{-/-}$ individuals exhibit a marked reduction in the number of teeth related to a failure of tooth replacement (G, H, and I, black arrowheads). (F and J) Schematic representations of the tooth phenotype of *scpp5*$^{+/+}$ and *scpp5*$^{-/-}$ fish, respectively (ventral tooth row, blue; medio-dorsal tooth row, orange; dorsal tooth row, green; and the dotted lines indicate missing tooth structures). Scale bars, 0.5 mm; a, anterior; p, posterior.

Putative sex determination locus

To identify the putative sex chromosome and sex locus in the common seadragon, whole-genome sequences of the male and female were compared by mapping the Illumina reads of male to female genomes and vice versa. A ~47-kb region on Chr4 (7523 to 7570 kb) was identified in the male that was completely absent in female. The coverage of male reads in this region is about half (~40×) of the average for Chr4 (~80×), suggesting that this might be a male-specific locus reflecting a XX/XY sex determination system. Hi-C interaction analyses also provided support to this hypothesis (Fig. 3A). This male-specific region is rich in repetitive sequences (~68%) (Fig. 3B), resembling the sex-determining locus on the Y chromosome of mammals, *Drosophila*, and some plants, and may play a role in the suppression of recombination and initiation of Y chromosome degeneration (*27*). Manual annotation and curation identified a single gene, the *anti-Müllerian hormone type II receptor* gene *amhr2* (designated as *amhr2y*) in this region [Chr4 7,541,024 to 7,555,173 base pairs (bp), Gene ID: EVM0015262]. Another *amhr2* gene, located on Chr10 (15,635,534 to 15,644,491 bp; Gene ID: EVM0017093) and common to both male and female genomes, was also found. Transcriptomic analysis showed that the canonical *amhr2* was expressed in both testis and ovary, with a higher expression level in the testis, whereas *amhr2y* expression was testis-specific (Fig. 3C), suggesting its importance for testis development and maintenance. The coding region of the *amhr2y* gene showed 80.32% nucleotide identity with that of the canonical *amhr2*, and their amino acids shared only 74.53% identity with a considerable divergence in the N-terminal. The *amhr2* gene contains 11 exons, while *amhr2y* has only 10 with the second exon missing. The first exon shows relatively lower similarity (~31%) between these two genes (Fig. 3F). The male alligator pipefish genome-sequenced in this study also contained *amhr2y* (Gene ID: EVM0001604) in its male-specific region, but on Chr21 (823 to 895 kb), a location different from that in the common seadragon (Fig. 3E and fig. S18A), suggesting a translocation event in this species. The alligator pipefish *amhr2* and *amhr2y* expression patterns were similar to those in the common seadragon (fig. S18B). Searches of previously sequenced male syngnathid genomes (*12*, *28*–*31*) showed that they all contain only the canonical *amhr2*.

Among teleosts, a Y-duplicated *amhr2* gene has previously been characterized as the master sex–determining gene in the yellow perch (*Perca flavescens*) (*32*). Phylogenetic analysis revealed that the *amhr2y* of common seadragon and alligator pipefish originated from a duplication event in their most recent common ancestor, whereas the yellow perch *amhr2y* evolved from an independent duplication event in its lineage (Fig. 3D). Unexpectedly, we found that the first three exons and introns of the alligator pipefish *amhr2y* are highly

similar (~80%) to the CA domain of its *B-cadherin–like* gene (Gene ID: EVM0014966), while they are only 42% identical to the corresponding region of its canonical *amhr2* gene. It appears that the *amhr2y* in the alligator pipefish underwent a gene fusion event involving its *B-cadherin–like* gene locus (Fig. 3, F and H).

The male-specific *amhr2* gene has previously been characterized as the master sex–determining gene in the yellow perch (*32*) and some pufferfish (*32*, *33*). In addition to *amhr2*, duplicate *amh* genes with independent origins in the Patagonian pejerrey (*Odontesthes hatcheri*), Nile tilapia (*Oreochromis niloticus*), and northern pike (*Esox lucius*) have been identified as their master sex–determining genes (*34*–*36*), highlighting the importance of the anti-Müllerian hormone (AMH) signaling pathway in the sex determination of teleosts (fig. S19). The Ser/Thr kinase domain in Amhr2 is critical for AMH signaling and male determination (*37*). In both common seadragon and alligator pipefish, Amhr2y proteins were predicted to have cytoplasmic Ser/Thr kinase domains with overall structure similar to that of Amhr2 (figs. S20 and S21A). However, the estimated ω (dN/dS ratio) values of the branch leading to the common ancestor of common seadragon and alligator pipefish *amhr2y* is 14.27, which is much higher than any other branch of *amhr2* (from 0.13 to 0.58), indicating an elevated evolutionary rate of the Ser/Thr kinase domain of the *amhr2y* gene (Fig. 3G). The Y-specific *amhr2* genes in common seadragon, alligator pipefish, and yellow perch were found to be under positive selection compared to their canonical *amhr2*. Furthermore, substitutions were found in one convergent amino acid site in the kinase domain, changing the Asp to Asn in the common seadragon and alligator pipefish and to Ala in the yellow perch (fig. S21B). In addition, Amhr2y proteins show shorter and variable extracellular regions compared to Amhr2 in all these three species (fig. S21C), which might be involved in ligand binding specificity. In conclusion, our results revealed a possible XX/XY-determining system and a putative sex determination gene in the common seadragon and alligator pipefish.

Loss of *scpp5* gene

The common seadragon, like other syngnathids, lacks oral and pharyngeal teeth. The *secretory calcium-binding phosphoprotein* (*scpp*) gene family was reported to be essential for teeth formation (*28*). *Scpp* genes fall into two subclasses: the acidic and P/Q-rich *scpps*. Acidic *scpps* are involved in the mineralization of bone and dentin, whereas P/Q-rich *scpps* are implicated in the formation of enamel or enameloid (*38*). Pseudogenization of the P/Q-rich *scpp* genes could lead to edentulous in many vertebrates such as birds, turtles, and some mammals (*39*). Previous studies have hypothesized that the loss of majority of P/Q-rich *scpp* genes might be responsible for edentulism in syngnathids (*28*, *31*). The genomes of the common seadragon and alligator pipefish contain only one intact P/Q-rich *scpp* gene, *fa93e10*, like the other sequenced syngnathid species (*31*). In addition, they contain only three of the 10 exons of *scpp5*, suggesting that this gene became pseudogenized in their common ancestor (fig. S22). In cichlid fish, *scpp5* was found to show higher expression in the lower pharyngeal jaws of large-toothed species compared to the small-toothed ones (*40*), suggesting that this P/Q-rich *scpp* gene might play a role in pharyngeal teeth formation. To verify this, we generated two zebrafish *scpp5* mutant lines using CRISPR-Cas9 technology. The pharyngeal teeth in the *scpp5*$^{-/-}$ zebrafish exhibited a distinct phenotype, particularly a decrease in numbers of functional teeth. The presence of pits in the jaw bones that is characteristic for previously attached teeth nevertheless indicate the former presence of teeth. This suggests that replacement in these positions is either delayed or arrested. In addition, there is a clear antero-posterior gradient of the loss of teeth (Fig. 4).

MATERIALS AND METHODS

Samples and nucleic acid preps

A total of six animals—four common seadragon (*P. taeniolatus*) and two alligator pipefish (*S. biaculeatus*) specimens were used for the study. The common seadragon specimens were donated by LanHai Co., a live animal supplier to aquariums in Guangzhou, to the Marine Biodiversity Collections of South China Sea, Chinese Academy of Science. The alligator pipefish samples were collected in Sanya (Hainan, P.R. China) and Xiamen (Fujian, P.R. China) from 2017 to 2019. Two common seadragon (one male and one female) and two alligator pipefish (one male and one female) individuals were used for whole-genome sequencing, and two other male common seadragon individuals were used for genome resequencing (table S1). Genomic DNA of each sample was extracted using a standard phenol-chloroform protocol. All experimental protocols in this study were approved by the Animal Experimental Ethical Inspection of Laboratory Animal Center, Huazhong Agricultural University, Wuhan, China. All efforts were made to minimize the suffering of the animals.

Genome sequencing

Two paired-end libraries with 350- or 270-bp insert size for each sample were constructed. Sequencing of each library was performed using the Illumina HiSeq 2500 system following the standard manufacturer's protocol (San Diego, USA) (table S2). The reads from the Illumina platform were quality filtered using the following strategy: (i) Reads from short-insert libraries were trimmed of four low-quality bases at both ends, while reads from long-insert libraries were trimmed of three low-quality bases; (ii) for long-insert libraries, duplicated reads were filtered out; (iii) reads with 10 or more Ns were discarded; and (iv) reads with more than 10-bp aligning to adapter sequences were filtered out. The quality-filtered reads were used for genome size estimation by the K-mer method (fig. S1 and table S3). For PacBio sequencing, genomic DNA of each sample was sheared to an average size of 20 kb using a g-TUBE device (Covaris, Woburn, MA, USA). The sheared DNA was purified and end repaired using polishing enzymes and then a blunt end ligation reaction followed by exonuclease treatment was applied to create a SMRTbell template according to the PacBio 20-kb template preparation protocol. A BluePippin device (Sage Science, Beverly, USA) was used to size-select the SMRTbell template and enrich large (>10 kbp) fragments. Single-molecule sequencing was then conducted on PacBio Sequel platform to generate long-read data. For Nanopore sequencing, libraries were constructed and sequenced on R9.4 Flow Cells using the MinION sequencer (ONT, UK) (table S4). In addition, paired-end sequencing libraries with an insert size of 350 bp were prepared for resequencing from two other male common seadragon individuals (Sample ID: Phtae_M2 and Phtae_M3). These libraries were sequenced using an Illumina NovaSeq 6000 platform. All sequencing was performed by BioMarker Technologies Company (Beijing, China).

Transcriptome sequencing

For common seadragon, brain, eye, gill, liver, heart, kidney, mesentery, intestine, muscle, fin, skin, leaf-like appendages, testis, and ovary

were collected. Total RNA was extracted from each tissue using a TRIzol kit (Life Technologies, Carlsbad, USA). The mRNA fractions were isolated from the total RNA extracts with the MicroPoly (A) Purist kit (Ambion, TX, USA). cDNA libraries were prepared for each tissue with the RNA-seq Library kit (Gnomegen, San Diego, CA, USA) following the manufacturer's instructions. Each paired-end cDNA library was sequenced with a read length of 150 bp using the Illumina HiSeq 2500 sequencing platform (table S5). In alligator pipefish, total RNA was extracted from gonads of three males (Sample IDs: Sybia_M2, Sybia_M3, and Sybia_M4) and three females (Sample IDs: Sybia_F1, Sybia_F2, and Sybia_F3). RNA was extracted and sequenced as described above to study the sex-related genes (table S6).

Hi-C sequencing

Hi-C libraries were prepared from the female common seadragon (Phtae_F1), the male seadragon (Phtae_M1), and the male alligator pipefish (Sybia_M1) at BioMarker Technologies Company (Beijing, China) and Frasergen Bioinformatics Company (Wuhan, China) as described previously (*41*). Briefly, blood sample was collected from each sample and spun down, and the cell pellet was resuspended and fixed in formaldehyde solution. DNA of each sample was isolated and the fixed chromatin was digested with the restriction enzyme DpnII overnight. The cohesive ends were labeled with Biotin-14-DCTP using Klenow enzyme and then religated with T4 DNA ligation enzyme. Subsequent DNA were sheared by sonication to the mean size of 350 bp. Hi-C libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adaptors. Biotin-containing fragments were isolated using streptavidin beads. All libraries were quantified by Qubit2.0, and insert size was checked using an Agilent 2100 and then quantified by quantitative polymerase chain reaction (PCR). Hi-C sequencing was performed by Illumina HiSeq 2500 platform, using paired-end of 150-bp reads. The Hi-C data were mapped to PacBio-based contigs using BWA (version 0.7.10-r789; mapping method: aln). Uniquely mapped data were used for chromosome-level scaffolding. HiC-Pro (version 2.8.1) was used for duplicate removal and quality controls and the remaining reads were valid interaction pairs for further assembly.

Genome assembly

PacBio subreads were corrected and trimmed using Canu (v1.5, available at https://github.com/marbl/canu, v1.5). The Canu first selects longer seed reads with the settings "genomeSize = 1000000000" and "corOutCoverage = 50," then detects overlaps through a high sensitive overlapper MHAP (mhap-2.1.2, option "corMhapSensitivity = low/normal/high"), and lastly performs an error correction through falcon_sense method (option "correctedErrorRate = 0.025"). In the assembly step, wtdbg generated a draft assembly with the command "wtdbg -i pbreads.fasta -t 40 -H -k 21 -S 1.02 -e 3 -o wtdbg" using error-corrected reads from Canu. A consensus assembly is obtained with the command "wtdbg-cns -t 40 -i wtdbg.ctg.lay -o wtdbg.ctg.lay.fa -k 15." Then, Illumina paired-end reads were also aligned to consensus assembly using BWA (version 0.7.10-r789; mapping method: MEM), and polishing step was performed with Pilon (version 1.22) software with parameters --mindepth 10 --changes --threads 4 --fix bases. The polishing step was iterated two times. Using contigs assembled from PacBio data, Hi-C data were used to correct misjoins in contigs, order, and orient contigs. Preassembly was performed for contig correction through splitting contigs into segments with an average length of 300 kb, and then the segments were preassembled with Hi-C data. Misassembled points were defined and broken when split segments could not be placed to the original position. Then, the corrected contigs were assembled using LACHESIS with parameters CLUSTER_MIN_RE_SITES = 225, CLUSTER_MAX_LINK_DENSITY = 2; ORDER_MIN_N_RES_IN_TRUN = 105; ORDER_MIN_N_RES_IN_SHREDS = 105 with Hi-C valid pairs. Gaps between ordered contigs were filled with 100 "N"s.

Genome annotation

De novo identification of repeats and TEs were performed using the PILER-DF (*42*) and the RepeatScout (*43*) under default parameters, filtered out short (<100 bp) sequences and those containing gaps (Ns > 5%), and then combined the results to obtain a consensus library. LTR_FINDER was used to identify LTR retrotransposons, and "einverted" software (EMBOSS) was used to detect terminal inverted repeats (TIRs). RepeatMasker and RepeatProteinMask (version 3.3.0, www.repeatmasker.org/) were used to identify TEs based on homology searches against the Repbase library (release 16.03, www.girinst.org/repbase/) using the parameters "-nolow -no_is -norna -parallel 1" and "-noLowSimple –pvalue 1e-4." An ab initio TE library was constructed by integrating two repeat finding programs RECON with RepeatModeler (version 1.08, www.repeatmasker.org/RepeatModeler.html). Using the repeat library constructed by RepeatModeler, we estimated the repeat content of the common seadragon and alligator pipefish genomes using RepeatMasker with the sensitive mode (-s) option. Ab initio gene prediction was performed using three programs Augustus, GlimmerHMM, and SNAP. GeMoMa program was run for homology-based prediction by aligning the assembled genome against those of stickleback, *Gasterosteus aculeatus*; platyfish, *Xiphophorus maculatus*; zebrafish, *Danio rerio*; tiger tail seahorse, *Hippocampus comes*; and lined seahorse, *Hippocampus erectus*. Then, the transcriptome data of the lined seahorse from our previous work (*28*) was downloaded and mapped onto genome, and gene prediction was performed by TransDecoder (http://transdecoder.github.io) and GeneMarkS-T. PASA was used to predict the Unigene sequences without reference assembly on the basis of transcriptome data. Last, EVidenceModeler (EVM) was used to integrate the prediction results obtained by the above three methods. Gene functions were annotated by searching publicly available databases including NR, KOG, Kyoto Encyclopedia of Genes and Genomes (KEGG), and TrEMBL. Then, the Gene Ontology (GO) annotation was retrieved by BLAST2GO (*44*). The gene pathways were assigned on the basis of the bidirectional best hit method using the KEGG Automatic Annotation Server. The motifs and domains of each gene model were predicted by InterProScan against PROSITE, HAMAP, Pfam, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, CATH-Gene3D, and PANTHER. For noncoding RNA annotation, microRNAs and ribosomal RNAs were predicted using Infernal with Rfam (*45*), while transfer RNAs were screened using tRNAscan-SE (*46*). For the pseudogene prediction, GenBlastA (v1.0.4) (*47*) was applied to identify the candidate pseudogene by homologous searching against genome data, and then GeneWise (v2.4.1) (*48*) was performed to search for immature termination and frameshift mutation of pseudogene.

Expansion and contraction of gene families

We compared the protein sequences of common seadragon to other 12 ray-finned fishes, including zebrafish, *D. rerio*; stickleback, *G. aculeatus*; Nile tilapia, *O. niloticus*; medaka, *Oryzias latipes*; fugu,

Takifugu rubripes; great blue-spotted mudskipper, *Boleophthalmus pectinirostris*; channel catfish, *Ictalurus punctatus*; spotted gar, *Lepisosteus oculatus*; tiger tail seahorse, *H. comes*; lined seahorse, *H. erectus*; gulf pipefish, *Syngnathus scovelli*; and Manado pipefish, *Microphis manadensis*, to identify potential homologs using BLASTP with an *E*-value of 1×10^{-10}. An in-house software solar (v 0.9.6) was used to refine the raw Blast results and conjoined the high-scoring segment pairs. Then, we evaluated the similarity between protein sequences using bit-score, and protein sequences were clustered into gene families using hcluster_sg, a hierarchical clustering algorithm in the Treefam pipeline (v 0.50) with the parameters: "-w 5 -s 0.33 -m 100,000." The male common seadragon genome was used for the expansion and contraction analysis. Expansion and contraction in gene family were calculated by the CAFÉ program (v 3.1) based on the birth-and-death model (*49*). The parameters "-p 0.01, -r 10000, -s" were set to search the birth and death parameter (λ) of genes based on a Monte Carlo resampling procedure, and birth and death parameters in gene families with the *P* value ≤ 0.01 have been reported. The gene families without homology in the SWISS-PROT database were filtered out to reduce the potential false-positive expansions or contractions caused by gene prediction. Then, gene families containing sequences that have multiple functional annotations were also removed.

The TE expansion analysis

The TE expansion history was constructed in the male common seadragon (Sample ID: Phtae_M1) by first recalculating the divergence of the identified TE copies in the genome with the corresponding consensus sequence in the TE library using Kimura distance and then estimating the percentage of TEs in the genome at different divergence levels (fig. S6). Since the Tc1/Mariner is the dominant type of TEs in common seadragon, the evolution of Tc1/Mariner was also analyzed as a representative. Other four syngnathid species were included in this comparison: alligator pipefish, *S. biaculeatus*; tiger tail seahorse, *H. comes*; lined seahorse, *H. erectus*; and Manado pipefish, *M. manadensis*. Using known Tc1/Mariner transposase proteins, blast searched was performed using an *E*-value $<1 \times 10^{-10}$ to identify transposase domains. Only protein sequences longer than 300 amino acids were selected for further analysis. The identified transposase sequences were aligned using MUSCLE and Tc1/Mariner phylogeny was generated using RAxML. The common characteristics of the Tc1/Mariner superfamily include TIRs, a DNA binding domain (includes a single HTH motif) and a catalytic domain (DDE-aspartate/aspartate/glutamate or DDD-aspartate/aspartate/aspartate) (*50*). The clustering of the Tc1/Mariner sequences from the five species was defined on the basis of the presence of DDE or DDD motifs within the catalytic domain. Phylogenetic clustering of Tc1/Mariner sequences were dated by calculating the average nucleotide divergence of individual copies to their ancestral consensus sequence (Ks) and estimating the substitutions per synonymous site per year (λ) following the formula: $T = Ks/\lambda$, where *T* represents the divergence time. The consensus sequences for a specific group were generated using the online Consensus Maker tool (www.hiv.lanl.gov/content/sequence/CONSENSUS/AdvCon.html). The Ks values were computed using KaKs_Calculator2.0 toolkit (*51*).

Phylogenomic analysis

To determine the phylogenetic position of common seadragon and alligator pipefish with respect to the other syngnathid fishes, we performed phylogenomic analysis using whole-genome protein datasets from several representative ray-finned fishes. Protein datasets were obtained from Ensembl-FTP release-96 (fugu *T. rubripes*; stickleback, *G. aculeatus*; medaka, *O. latipes*; Nile tilapia, *O. niloticus*; turbot, *Scophthalmus maximus*; tongue sole, *Cynoglossus semilaevis*; greater amberjack, *Seriola dumerili*; giant fin mudskipper, *Periophthalmus magnuspinnatus*; zebrafish, *D. rerio*; and spotted gar, *L. oculatus*) or other sources—Pacific bluefin tuna, *Thunnus orientalis* (*52*); Atlantic cod, *Gadus morhua* (gadMor2) (*53*); Asian arowana, *Scleropages formosus* (*54*); tiger tail seahorse, *H. comes* (*28*); lined seahorse, *H. erectus* (*30*); gulf pipefish, *S. scovelli* (*29*); Manado pipefish, *M. manadensis* (*31*); common seadragon, *P. taeniolatus*; and alligator pipefish, *S. biaculeatus* (present study). OrthoFinder version 2.2.7 (*55*) was used at default settings to identify one-to-one orthologs from the 19 ray-finned fish species. Multiple alignments were generated for each of the orthologous groups using MAFFT (v 7.475, https://mafft.cbrc.jp/alignment/software/). Gblocks was used to remove poorly aligned regions for each of the orthologous groups with the "allowed gap positions" set to "With Half." After alignment and trimming, all one-to-one orthologous genes were concatenated into a single supergene. ProteinModelSelection.pl provided with RAxML (v 8.2.12) (*56*) was used to deduce the best-suited substitution model for the concatenated alignment. A maximum likelihood (ML) tree was generated using RAxML. For the ML analysis, we used the best fit substitution model as deduced by ProteinModelSelection.pl and 1000 replicates for bootstrap support.

Molecular dating was performed using MCMCtree (v 4.9j) contained in PAML software. The time calibration was based on a total of eight calibration points including three dated fossil records: (i)†*Acentrophorus varians* (~254 to 260 Ma ago), the oldest known neopterygian (*12*); (ii)†*Plectocretacicus clarae* (~97 to 153 Ma ago), earliest stem tetraodontiform (*57*); and (iii)†*Prosolenostomus lessenii* (~48 to 50 Ma ago), the oldest syngnathid fossil (*58*, *59*). The 95% HPD intervals of estimated node ages were shown at each node (fig. S7).

Histological analysis of the leaf-like appendages

Samples of fins and leaf-like appendages were fixed in 4% paraformaldehyde (PFA) at 4°C overnight, dehydrated via a series of graded ethanol concentrations, and embedded in paraffin blocks. Sections at 5-μm thickness were made and then stained by hematoxylin and eosin and Masson's trichrome (fig. S8).

Transcriptome analysis

Raw data (raw reads) of FASTQ format from different tissues (brain, eye, gill, liver, intestine, muscle, fin, skin, and leaf-like appendages) were first processed using in-house Perl scripts. Clean reads were mapped to the male common seadragon genome using the TopHat2 program (*60*). Gene expression levels were estimated with FPKM values using the Cufflinks program. Principal components analysis was conducted to estimate the correlation across leaf-like appendages from different position on the body. Four paired (XT01, XT09, XT11, and XT12) and four unpaired leaf-like appendages (XT02, XT10, XT42, and XT43) were used in the comparison. Statistically significant differences between paired and unpaired leaf-like appendages were compared by *t* test using SPSS (v 19.0) software. Since there was no significant difference between paired and unpaired leaf-like appendages, three samples (XT02, XT09, and XT12) were randomly selected for gene recruitment and differential expression analysis, respectively. High-expressed genes in leaf-like appendages were defined as

FPKM value > 100. Furthermore, KOBAS (*61*) software was used to test the statistical enrichment of genes in KEGG pathways, and *P* values < 0.05 were considered to be significantly enriched pathways. To obtain a general gene expression profile of leaf-like appendages, transcriptomes from nine tissues (brain, eye, gill, liver, intestine, muscle, fin, skin, and leaf-like appendage) were used to determine expression patterns. To identify genes with tissue-specific expression, we calculated tissue specificity indices as follows. First, we filtered out weakly expressed genes, for which the maximum detected expression level in all tissues was less than 1 FPKM. We then used the remaining genes to calculate τ tissue-specific expression indices, which are the best indicators according to previous studies (*20*, *62*). The values of τ range between 0 and 1. Values close to 1 indicate completely tissue-specific expressed genes, and values close to 0 indicate ubiquitously expressed genes. Genes specifically expressed in leaf-like appendages were defined as those that have a τ index exceeding 0.8 and are expressed most strongly or second most strongly in leaf-like appendages. The DEGs for leaf-like appendage and fin were checked using DEGSeq2 and identified on the basis of corrected *P* values (Q-value) and false discovery rates. Last, we identified 2375 DEGs with a $|\log_2(\text{fold change})| > 1$ and corrected *P* value < 0.05 (using the Benjamini-Hochberg algorithm). As for specific expressed genes (yes or no) in leaf-like appendages and fins, methods were defined as ($A \geq 0.5$; $B = 0$) and ($A = 0$; $B \geq 0.5$) based on FPKM values.

Expansion of *MHC I* genes

The male genomes of common seadragon and alligator pipefish (Sample ID: Phtae_M1 and Sybia_M1) were used in the analysis of *MHC I* genes, and the genome of tiger tail seahorse, lined seahorse, gulf pipefish and Manado pipefish were included for comparison. We searched the *MHC I* genes in these species by both TBLASTN and checking the gene annotation. In total, 32 and 14 *MHC I* genes were found in common seadragon and alligator pipefish, respectively. Nineteen *MHC I* genes in the common seadragon genome were found to be tandemly organized into a cluster, and the RTE and Gypsy TEs are quite abundant in this region, although Tc1-Mariner is the dominating type of TE in the whole genome (fig. S15).

We examined the relatedness of *MHC I* genes in these six species. Phylogenetic tree showed that the *MHC I* genes group into two divergent clades, and the expansion of common seadragon *MHC I* genes was mainly clustered in clade I (fig. S16A). The gene expression pattern of each *MHC I* gene was estimated by transcriptomic analysis in common seadragon. Among the 12 tissues (leaf-like appendages, skin, gill, fin, intestine, testis, ovary, liver, kidney, muscle, eye, and brain), most *MHC I* genes, especially the tandemly repeated ones, have high expression levels in leaf-like appendages, skin, gill, fin, and intestine (fig. S16B).

PCR validation of *amhr2* and *amhr2y* genes

Degenerate primers were designed for both *amhr2* and *amhr2y* in common seadragon and alligator pipefish to verify the male-specificity of the *amhr2y* (table S22). The PCR mix composed of 12.5 μl of PCR Master Mix (Qiagen, Germany), 0.5 μl of each primer, and 2 μl of genomic DNA with the concentrations of ~100 ng/μl, and ddH_20 was added to a total reaction volume of 25 μl. Thermocycling conditions were as follows: first cycle of 5 min at 95°C, then 35 cycles of 30 s at 95°C, 30 s at 57°C and 120 s at 72°C, and followed by 10 min of incubation at 72°C. PCR products were loaded on a 1.5% agarose gel and run at 130 V for 25 min. Results were photographed under ultraviolet light.

Identification of the sex-specific region

The absolute and relative read depths for the male and female along the male genome were conducted to look for the male-specific regions in common seadragon and alligator pipefish, respectively. *psass* (v 2.0.0: https://zenodo.org/record/2615936#.XTyIS3s6_AI) was used for this analysis with default parameters except --window-size, which was set to 1k.

Phylogenetic analysis of *amhr2* and *amhr2y*

A total of 71 vertebrates containing 39 fish species from 21 orders were included in the analysis (table S23). Amino acid sequences of each gene were aligned using Clustal W with minor manual adjustments. Bayesian phylogenetic analysis was performed using MrBayes (v 3.2.7) (*63*). Trees and parameters were sampled every 1000 generations over a total of 1,000,000 generations, with the first 25% of the samples discarded as burn-in. The average SD of split frequencies is 0.006244. A majority consensus tree with Bayesian posterior probabilities was computed. Inferred tree topologies were rendered in FigTree (v 1.4.4) software.

Structural features of Amhr2 and Amhr2y

Amino acid sequences of Amhr2 and Amhr2y in common seadragon and alligator pipefish were aligned together with the Amhr2a and Amhr2by of yellow perch (*P. flavescens*) using the Clustal W. Transmembrane region and Ser/Thr kinase domain were predicted by InterPro (www.ebi.ac.uk/interpro/) (fig. S20 and S21). The three-dimensional structures of the Ser/Thr kinase domains of Amhr2 and Amhr2y of common seadragon and alligator pipefish were predicted using I-TASSER (*64*). The rainbow color is from N (red) to C (blue) termini (fig. S21A).

Evolutionary rate of *amhr2* and *amhr2y*

DNA sequences of Ser/Thr kinase domain in the *amhr2* and *amhr2y* of six Syngnathidae species were aligned using Clustal W followed by minor manual adjustments. The optimal model of sequence evolution was determined by jModelTest 3.7, and phylogenetic tree was inferred using Bayesian methods by MrBayes (v 3.1.2) (*63*). EasyCodeML program in PAML (v 4.9) (*65*) was used to estimate the ratio (ω) of the rate of nonsynonymous to synonymous substitutions (dN/dS) under the branch model by comparing the M0 model (all genes evolved with the same rates, model = 0) with an M1 model (free ratio, allows the ω ratio to vary on each branch, model = 1). M1 model fits better with a likelihood ratio test *P* value of 9.04×10^{-9}. In addition, ω value was denoted at right to branches, $\omega > 1$, suggests positive selection; whereas $\omega \approx 1$ indicates neutral evolution and $\omega < 1$ indicates purifying selection (Fig. 3G).

CRISPR-Cas9–based knockout of *scpp5*

To investigate the phenotypic consequences of *scpp5* loss in the common seadragon, the CRISPR-Cas9 strategy was used to generate a *scpp5* mutant zebrafish line. Zebrafish were kept at 26° to 28°C under a controlled light cycle (14 hours light, 10 hours dark) to induce spawning. The offspring were used for subsequent experiments. The guide RNAs (gRNAs) were designed targeting zebrafish *scpp5* in the exon 5 (table S25) according to the study of Moreno-Mateos *et al.* (*66*). The gRNA template DNA for the in vitro transcription was made

using the following overlap PCR system and then run at 98°C for 30 s, 45 cycles of (98°C for 10 s, 60°C for 10 s, and 72°C for 15 s), 72°C for 5 min, 10°C ∞.

The generated gRNA template and synthesized Cas9 plasmid (pCS2-nCas9n) DNA were then used for the in vitro transcription using the mMessage mMachine T7 Transcription Kit (Thermo Fisher Scientific) and purified using the RNA cleanup protocol from the RNAeasy mini kit (Qiagen, Germany). Purified gRNAs (~80 ng/μl) were coinjected with Cas9 mRNA (~400 ng/μl) into zebrafish embryos (F0 fish) at the one-cell stage. These F0 fish were raised to maturity and genotyped using fin clipping. The corresponding primers (forward: scpp5_F: 5′-AAAACAAGCCAACCAG-3′; reverse: scpp5_R: 5′-TTTCTGCCGGGAGCT-3′) were used to screen out founders with site mutations. The adult founders were outcrossed with wild-type fish to obtain F1 fish, which were subsequently genotyped and outcrossed with wild-type fish to obtain F2 fish. Then, heterozygous F2 individuals were intercrossed to obtain homozygous F3 fish. Next, mineralized bone was stained with Alizarin Red using a modified protocol from the lab of P. Eckhard Witten in Ghent University (Belgium) for larvae and adult fish: (i) fix the specimens in 4% PFA overnight; (ii) wash specimens three times with phosphate buffered saline with Tween 20 (PBST) (0.1% Tween 20), 5min each time; (iii) prepare the bleach solution: mix 3% H_2O_2 and 2% KOH according to 1:2; (iv) add bleach solution to the specimens and process the sample for 10 min (depending on the size of the specimens, do not take too long, carefully observe) until the specimens is transparent; (v) wash specimens three times with water, 5 min each time; (vi) transfer the sample to alizarin red staining solution (0.1%) for 1 to 2 hours (depending on the size of the specimen, timing observation, adjusting the dyeing time); (vii) wash three times with water; (viii) place the sample in 50% glycerol-KOH to make them transparent, and then transfer to 70% and store it in 100% glycerol for examination. (ix) The stained specimens were analyzed and photographed under an Olympus SZX2 Stereo Microscope (Japan).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/34/eabg5196/DC1

REFERENCES AND NOTES

1. C. E. Dawson, *Indo-Pacific Pipefishes Red Sea to the Americas* (Gulf Coast Research Laboratory, Ocean Springs, MS, 39564, 1985).
2. C. M. Whittington, C. R. Friesen, The evolution and physiology of male pregnancy in syngnathid fishes. *Biol. Rev.* **95**, 1252–1272 (2020).
3. C. Li, Y. Li, G. Qin, Z. Chen, M. Qu, B. Zhang, X. Han, X. Wang, P. Y. Qian, Q. Lin, Regulatory role of retinoic acid in male pregnancy of the seahorse. *Innovation* **1**, 100052 (2020).
4. A. B. Wilson, A. Vincent, I. Ahnesjö, A. Meyer, Male pregnancy in seahorses and pipefishes (family Syngnathidae): Rapid diversification of paternal brood pouch morphology inferred from a molecular phylogeny. *J. Hered.* **92**, 159–166 (2001).
5. H. Hamilton, N. Saarman, G. Short, A. B. Sellas, B. Moore, T. Hoang, C. L. Grace, M. Gomon, K. Crow, W. Brian Simison, Molecular phylogeny and patterns of diversification in syngnathid fishes. *Mol. Phylogenet. Evol.* **107**, 388–403 (2017).
6. J. Sanchez Camara, D. J. Booth, X. Turon, Reproductive cycle and growth of *Phyllopteryx taeniolatus*. *J. Fish Biol.* **67**, 133–148 (2005).
7. A. B. Wilson, I. Ahnesjo, A. C. J. Vincent, A. Meyer, The dynamics of male brooding, mating patterns, and sex roles in pipefishes and seahorses (family Syngnathidae). *Evolution* **57**, 1374–1386 (2003).
8. D. Bachtrog, J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T. L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, Sex determination: Why so many ways of doing it? *PLoS Biol.* **12**, e1001899 (2014).
9. J. Sanchez Camara, D. J. Booth, J. Murdoch, D. Watts, X. Turon, Density, habitat use and behaviour of the weedy seadragon *Phyllopteryx taeniolatus* (Teleostei:Syngnathidae) around Sydney, New South Wales, Australia. *Mar. Freshw. Res.* **57**, 737–745 (2006).
10. A. B. Wilson, J. W. Orr, The evolutionary origins of Syngnathidae: Pipefishes and seahorses. *J. Fish Biol.* **78**, 1603–1623 (2011).
11. S. J. Longo, B. C. Faircloth, A. Meyer, M. W. Westneat, M. E. Alfaro, P. C. Wainwright, Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. *Mol. Phylogenet. Evol.* **113**, 33–48 (2017).
12. O. Roth, M. H. Solbakken, O. K. Torresen, T. Bayer, M. Matschiner, H. T. Baalsrud, S. N. K. Hoff, M. S. O. Brieuc, D. Haase, R. Hanel, T. B. H. Reusch, S. Jentoft, Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9431–9439 (2020).
13. M. E. Alfaro, B. C. Faircloth, R. C. Harrington, L. Sorenson, M. Friedman, C. E. Thacker, C. H. Oliveros, D. Cerny, T. J. Near, Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. *Nat. Ecol. Evol.* **2**, 688–696 (2018).
14. T. J. Near, A. Dornburg, R. I. Eytan, B. P. Keck, W. L. Smith, K. L. Kuhn, J. A. Moore, S. A. Price, F. T. Burbrink, M. Friedman, P. C. Wainwright, Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12738–12743 (2013).
15. L. C. Hughes, G. Orti, Y. Huang, Y. Sun, C. C. Baldwin, A. W. Thompson, D. Arcila, R. R. Betancur, C. Li, L. Becker, N. Bellora, X. Zhao, X. Li, M. Wang, C. Fang, B. Xie, Z. Zhou, H. Huang, S. Chen, B. Venkatesh, Q. Shi, Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6249–6254 (2018).
16. J. Stiller, N. G. Wilson, G. W. Rouse, A spectacular new species of seadragon (Syngnathidae). *R. Soc. Open Sci.* **2**, 140458 (2015).
17. C. Bonnans, J. Chou, Z. Werb, Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).
18. L. H. Shen, Y. Li, Q. Gao, S. Savant Bhonsale, M. Chopp, Down-regulation of neurocan expression in reactive astrocytes promotes axonal regeneration and facilitates the neurorestorative effects of bone marrow stromal cells in the ischemic rat brain. *Glia* **56**, 1747–1754 (2008).
19. T. Tanji, A. Ohashi Kobayashi, S. Natori, Participation of a galactose-specific C-type lectin in *Drosophila* immunity. *Biochem. J.* **396**, 127–138 (2006).
20. Y. Wang, C. Zhang, N. Wang, Z. Li, R. Heller, R. Liu, Y. Zhao, J. Han, X. Pan, Z. Zheng, X. Dai, C. Chen, M. Dou, S. Peng, X. Chen, J. Liu, M. Li, K. Wang, C. Liu, Z. Lin, L. Chen, F. Hao, W. Zhu, C. Song, C. Zhao, C. Zheng, J. Wang, S. Hu, C. Li, H. Yang, L. Jiang, G. Li, M. Liu, T. S. Sonstegard, G. Zhang, Y. Jiang, W. Wang, Q. Qiu, Genetic basis of ruminant headgear and rapid antler regeneration. *Science* **364**, eaav6335 (2019).
21. M. A. Akimenko, S. L. Johnson, M. Westerfield, M. Ekker, Differential induction of four msx homeobox genes during fin development and regeneration in zebrafish. *Development* **121**, 347–357 (1995).
22. K. D. Poss, J. Shen, A. Nechiporuk, G. McMahon, B. Thisse, C. Thisse, M. T. Keating, Roles for Fgf signaling during zebrafish fin regeneration. *Dev. Biol.* **222**, 347–358 (2000).
23. R. Freitas, G. Zhang, M. J. Cohn, Evidence that mechanisms of fin development evolved in the midline of early vertebrates. *Nature* **442**, 1033–1037 (2006).
24. T. Nakamura, A. R. Gehrke, J. Lemberg, J. Szymaszek, N. H. Shubin, Digits and fin rays share common developmental histories. *Nature* **537**, 225–228 (2016).
25. C. Frantz, K. M. Stewart, V. M. Weaver, The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200 (2010).
26. C. Peng, J. L. Ren, C. Deng, D. Jiang, J. Wang, J. Qu, J. Chang, C. Yan, K. Jiang, R. W. Murphy, D. D. Wu, J. T. Li, The genome of Shaw's sea snake (*Hydrophis curtus*) reveals secondary adaptation to its marine environment. *Mol. Biol. Evol.* **37**, 1744–1760 (2020).
27. D. Bachtrog, Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
28. Q. Lin, S. Fan, Y. Zhang, M. Xu, H. Zhang, Y. Yang, A. P. Lee, J. M. Woltering, V. Ravi, H. M. Gunter, W. Luo, Z. Gao, Z. W. Lim, G. Qin, R. F. Schneider, X. Wang, P. Xiong, G. Li, K. Wang, J. Min, C. Zhang, Y. Qiu, J. Bai, W. He, C. Bian, X. Zhang, D. Shan, H. Qu, Y. Sun, Q. Gao, L. Huang, Q. Shi, A. Meyer, B. Venkatesh, The seahorse genome and the evolution of its specialized morphology. *Nature* **540**, 395–399 (2016).
29. C. M. Small, S. Bassham, J. Catchen, A. Amores, A. M. Fuiten, R. S. Brown, A. G. Jones, W. A. Cresko, The genome of the Gulf pipefish enables understanding of evolutionary innovations. *Genome Biol.* **17**, 258 (2016).
30. Q. Lin, Y. Qiu, R. Gu, M. Xu, J. Li, C. Bian, H. Zhang, G. Qin, Y. Zhang, W. Luo, J. Chen, X. You, M. Fan, M. Sun, P. Xu, B. Venkatesh, J. Xu, H. Fu, Q. Shi, Draft genome of the lined seahorse, *Hippocampus erectus*. *Gigascience* **6**, 1–6 (2017).
31. Y.-H. Zhang, V. Ravi, G. Qin, H. Dai, H.-X. Zhang, F.-M. Han, X. Wang, Y.-H. Liu, J.-P. Yin, L.-M. Huang, B. Venkatesh, Q. Lin, Comparative genomics reveal shared genomic changes in syngnathid fishes and signatures of genetic convergence with placental mammals. *Natl. Sci. Rev.* **7**, 964–977 (2020).
32. R. Feron, M. Zahm, C. Cabau, C. Klopp, C. Roques, O. Bouchez, C. Eche, S. Valiere, C. Donnadieu, P. Haffray, A. Bestin, R. Morvezen, H. Acloque, P. T. Euclide, M. Wen, E. Jouano, M. Schartl, J. H. Postlethwait, C. Schraidt, M. R. Christie, W. A. Larson, A. Herpin, Y. Guiguen, Characterization of a Y-specific duplication/insertion of the anti-Mullerian

hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Mol. Ecol. Resour.* **20**, 531–543 (2020).
33. T. Kamiya, W. Kai, S. Tasumi, A. Oka, T. Matsunaga, N. Mizuno, M. Fujita, H. Suetake, S. Suzuki, S. Hosoya, S. Tohari, S. Brenner, T. Miyadai, B. Venkatesh, Y. Suzuki, K. Kikuchi, A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLOS Genet.* **8**, e1002798 (2012).
34. R. S. Hattori, Y. Murai, M. Oura, S. Masuda, S. K. Majhi, T. Sakamoto, J. I. Fernandino, G. M. Somoza, M. Yokota, C. A. Strüssmann, A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2955–2959 (2012).
35. Q. Pan, R. Feron, A. Yano, R. Guyomard, E. Jouanno, E. Vigouroux, M. Wen, J. M. Busnel, J. Bobe, J. P. Concordet, H. Parrinello, L. Journot, C. Klopp, J. Lluch, C. Roques, J. Postlethwait, M. Schartl, A. Herpin, Y. Guiguen, Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLOS Genet.* **15**, e1008013 (2019).
36. M. Li, Y. Sun, J. Zhao, H. Shi, S. Zeng, K. Ye, D. Jiang, L. Zhou, L. Sun, W. Tao, Y. Nagahama, T. D. Kocher, D. Wang, A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in nile tilapia, oreochromis niloticus. *PLOS Genet.* **11**, e1005678 (2015).
37. C. Morinaga, D. Saito, S. Nakamura, T. Sasaki, S. Asakawa, N. Shimizu, H. Mitani, M. Furutani-Seiki, M. Tanaka, H. Kondoh, The hotei mutation of medaka in the anti-Mullerian hormone receptor causes the dysregulation of germ cell and sexual development. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9691–9696 (2007).
38. K. Kawasaki, K. M. Weiss, SCPP gene evolution and the dental mineralization continuum. *J. Dent. Res.* **87**, 520–531 (2008).
39. R. W. Meredith, G. Zhang, M. T. Gilbert, E. D. Jarvis, M. S. Springer, Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* **346**, 1254390 (2014).
40. N. Karagic, R. F. Schneider, A. Meyer, C. D. Hulsey, A genomic cluster containing novel and conserved genes is associated with cichlid fish dental developmental convergence. *Mol. Biol. Evol.* **37**, 3165–3174 (2020).
41. T. Xie, J. F. Zheng, S. Liu, C. Peng, Y. M. Zhou, Q. Y. Yang, H. Y. Zhang, De novo plant genome assembly based on chromatin interactions: A case study of *Arabidopsis thaliana*. *Mol. Plant* **8**, 489–492 (2015).
42. R. C. Edgar, E. W. Myers, PILER: Identification and classification of genomic repeats. *Bioinformatics* **21**, i152–i158 (2005).
43. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
44. A. Conesa, S. Götz, J. M. García Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
45. S. Griffiths Jones, S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, A. Bateman, Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
46. T. M. Lowe, S. R. Eddy, tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
47. R. She, J. S.-C. Chu, K. Wang, J. Pei, N. Chen, GenBlastA: Enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**, 143–149 (2009).
48. E. Birney, M. Clamp, R. Durbin, GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
49. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
50. R. H. A. Plasterk, Z. Izsvák, Z. Ivics, Resident aliens: The Tc1/mariner superfamily of transposable elements. *Trends Genet.* **15**, 326–332 (1999).
51. D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**, 77–80 (2010).
52. Y. Nakamura, K. Mori, K. Saitoh, K. Oshima, M. Mekuchi, T. Sugaya, Y. Shigenobu, N. Ojima, S. Muta, A. Fujiwara, M. Yasuike, I. Oohara, H. Hirakawa, V. S. Chowdhury, T. Kobayashi, K. Nakajima, M. Sano, T. Wada, K. Tashiro, K. Ikeo, M. Hattori, S. Kuhara, T. Gojobori, K. Inouye, Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11061–11066 (2013).
53. O. K. Tørresen, B. Star, S. Jentoft, W. B. Reinar, H. Grove, J. R. Miller, B. P. Walenz, J. Knight, J. M. Ekholm, P. Peluso, R. B. Edvardsen, A. Tooming-Klunderud, M. Skage, S. Lien, K. S. Jakobsen, A. J. Nederbragt, An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics* **18**, 95 (2017).
54. C. Bian, Y. Hu, V. Ravi, I. S. Kuznetsova, X. Shen, X. Mu, Y. Sun, X. You, J. Li, X. Li, Y. Qiu, B.-H. Tay, N. M. Thevasagayam, A. S. Komissarov, V. Trifonov, M. Kabilov, A. Tupikin, J. Luo, Y. Liu, H. Song, C. Liu, X. Wang, D. Gu, Y. Yang, W. Li, G. Polgar, G. Fan, P. Zeng, H. Zhang, Z. Xiong, Z. Tang, C. Peng, Z. Ruan, H. Yu, J. Chen, M. Fan, Y. Huang, M. Wang, X. Zhao, G. Hu, H. Yang, J. Wang, J. Wang, X. Xu, L. Song, G. Xu, P. Xu, J. Xu, S. J. O'Brien, L. Orbán, B. Venkatesh, Q. Shi, The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Sci. Rep.* **6**, 24501 (2016).
55. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
56. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
57. R. E. Broughton, R. Betancur-R, C. Li, G. Arratia, G. Ortí, Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLOS Currents* **5**, ecurrents.tol.2ca8041495ffafd8041490c8092756e75247483e, (2013).
58. P. R. Teske, L. B. Beheregaray, Evolution of seahorses' upright posture was linked to Oligocene expansion of seagrass habitats. *Biol. Lett.* **5**, 521–523 (2009).
59. A. F. Bannikov, G. Carnevale, Eocene ghost pipefishes (Teleostei, Solenostomidae) from Monte Bolca. *Italy. Boll. Soc. Paleontol. Ital.* **56**, 319–333 (2017).
60. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
61. X. Mao, T. Cai, J. G. Olyarchuk, L. Wei, Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793 (2005).
62. N. Kryuchkova Mostacci, M. Robinson Rechavi, A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.* **18**, 205–214 (2017).
63. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
64. Y. Zhang, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
65. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
66. M. A. Moreno-Mateos, C. E. Vejnar, J.-D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha, A. J. Giraldez, CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–988 (2015).

Acknowledgments: We acknowledge the Marine Biodiversity Collections of South China Sea, CAS for providing samples; the Laboratory Animal Center of Huazhong Agricultural University for providing experimental facilities; and the Agency for Science, Q. Zhou, and X. Lu for insightful comments and suggestions that improved the manuscript. **Funding:** This research was supported by Key Research Program of Frontier Sciences of CAS, ZDBS-LY-DQC004 (Q.L.) Strategic Priority Research Program of the Chinese Academy of Sciences, XDB42030204 (Q.L.) National Natural Science Foundation of China, 41825013 (Q.L.) Key Special Project for Introduced Talents Team of GML (Guangzhou), GML2019ZD0407 (Q.L.) K.C. Wong Education Foundation (Q.L.) Biomedical Research Council of A*STAR, Singapore (B.V.). **Author contributions:** Research design: Q.L., B.V., and A.M. Genome analyses: Y.Z., Z.C., H.Y., X.W., G.Q., Huixian Zhang, H.J., Z.Z., C.L., Hao Zhang, B.Z., J.Z., Y.W., L.T., L.M., F.W., H. Zhe., and J.Y. TE analysis: Y.Z. and H.Y. Phylogenomic analysis: V.R. and H.Y. Sex-specific region analysis: M.Q. Transcriptomic analysis: Y.L. Gene editing: S.W., P.E.W., A.H., and Z.G. Writing—original draft: M.Q., Y.L., and V.R. Writing—review and editing: B.V., A.M., and Q.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The whole-genome assemblies of the common seadragon and alligator pipefish have been submitted to NCBI under PRJNA669508. The raw reads of the RNA-seq are submitted to NCBI under the accession number PRJNA671520.

Submitted 11 January 2021
Accepted 1 July 2021
Published 18 August 2021
10.1126/sciadv.abg5196

Citation: M. Qu, Y. Liu, Y. Zhang, S. Wan, V. Ravi, G. Qin, H. Jiang, X. Wang, H. Zhang, B. Zhang, Z. Gao, A. Huysseune, Z. Zhang, H. Zhang, Z. Chen, H. Yu, Y. Wu, L. Tang, C. Li, J. Zhong, L. Ma, F. Wang, H. Zheng, J. Yin, P. E. Witten, A. Meyer, B. Venkatesh, Q. Lin, Seadragon genome analysis provides insights into its phenotype and sex determination locus. *Sci. Adv.* **7**, eabg5196 (2021).

Science Advances

Seadragon genome analysis provides insights into its phenotype and sex determination locus

Meng Qu, Yali Liu, Yanhong Zhang, Shiming Wan, Vydianathan Ravi, Geng Qin, Han Jiang, Xin Wang, Huixian Zhang, Bo Zhang, Zexia Gao, Ann Huysseune, Zhixin Zhang, Hao Zhang, Zelin Chen, Haiyan Yu, Yongli Wu, Lu Tang, Chunyan Li, Jia Zhong, Liming Ma, Fengling Wang, Hongkun Zheng, Jianping Yin, Paul Eckhard Witten, Axel Meyer, Byrappa Venkatesh and Qiang Lin

Sci Adv **7** (34), eabg5196.
DOI: 10.1126/sciadv.abg5196