communications biology

Article

https://doi.org/10.1038/s42003-024-06103-x

Convergent gene losses and pseudogenizations in multiple lineages of stomachless fishes

Akira Kato[1,2,3,4] ✉, Supriya Pipil[5], Chihiro Ota[1], Makoto Kusakabe[5,6], Taro Watanabe[5], Ayumi Nagashima[1], An-Ping Chen[4], Zinia Islam[2], Naoko Hayashi[2], Marty Kwok-Shing Wong[5,7], Masayuki Komada[1,8], Michael F. Romero[4,9] & Yoshio Takei[5]

The regressive evolution of independent lineages often results in convergent phenotypes. Several teleost groups display secondary loss of the stomach, and four gastric genes, *atp4a*, *atp4b*, *pgc*, and *pga2* have been co-deleted in agastric (stomachless) fish. Analyses of genotypic convergence among agastric fishes showed that four genes, *slc26a9*, *kcne2*, *cldn18a*, and *vsig1*, were co-deleted or pseudogenized in most agastric fishes of the four major groups. *kcne2* and *vsig1* were also deleted or pseudogenized in the agastric monotreme echidna and platypus, respectively. In the stomachs of sticklebacks, these genes are expressed in gastric gland cells or surface epithelial cells. An ohnolog of *cldn18* was retained in some agastric teleosts but exhibited an increased non-synonymous substitution when compared with gastric species. These results revealed novel convergent gene losses at multiple loci among the four major groups of agastric fish, as well as a single gene loss in the echidna and platypus.

Actinopterygii (ray-finned fishes) consists of 44 orders, 453 families, and approximately 30,000 species, thereby constituting the largest class of fishes, as well as greater than half of all extant vertebrates[1–3]. The stomach is absent from the gastrointestinal tract in certain Actinopterygii orders, while others have true stomachs that secrete gastric acid and pepsinogen from the gastric gland. Cypriniformes (~3,200 species; e.g., minnows), Beloniformes and Cyprinodontiformes (~1200 species; e.g., medaka and killifish, respectively), Tetraodontiformes (~3500 species; e.g., pufferfishes), and Labriformes (~600 species; e.g., wrasse) are the main predominantly agastric orders of this class[4,5]. These groups are phylogenetically scattered, showing that the Actinopterygii originally possessed a stomach, but this organ disappeared in the ancestors of each agastric lineage individually. In the most recent review of stomach loss in fishes, Wilson and Castro[5] estimated that 7% of families and 20–27% of fish species are agastric, and at least 15 individual stomach loss events have occurred in fishes during evolution.

Secondary loss of an organ or tissue is a type of regressive evolution that has received considerable attention as a model of evolution, development, and physiology. These losses are convergent phenotypes, suggesting the presence of a specific benefit and selection in each lineage. For example, secondary eye and pigment losses are observed in cave animals such as cavefishes (*Astyanax mexicanus*, *Amblyopsis rosae*, and *Typhlichthys subterraneus*) and cave salamanders[6,7], with eye loss suggested to relate to the conservation of metabolic energy[8]. Other examples are the loss of the swim bladder in Pleuronectiformes, Gobiiformes, and Scorpaenidae[9], and the disappearance of scales in some lineages of Actinopterygii[10]. Most stenohaline marine fishes lack a distal tubule from the nephron of the kidney, and glomeruli are absent in a small number of marine teleosts as they have minimal functional significance[11]. Snakes and scincid lizards have lost their limbs[12,13], most cetaceans present missing hind limbs[14], and the aquatic frog *Barbourula kalimantanensis* lacks a lung[15]. In the platypus, the stomach is completely aglandular and has been reduced to a simple dilatation of the lower esophagus[16]. In echidna, the small stomach contains a high gastric fluid pH but lacks a gastric gland[16]. The secondary stomach loss in Actinopterygii, as well as the loss of gastric gland in monotremes, is an

[1]School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan. [2]Department of Biological Sciences, Tokyo Institute of Technology, Yokohama, Japan. [3]Center for Biological Resources and Informatics, Tokyo Institute of Technology, Yokohama, Japan. [4]Department of Physiology & Biomedical Engineering, Mayo Clinic College of Medicine & Science, Rochester, MN, USA. [5]Department of Marine Bioscience, Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan. [6]Department of Biological Sciences, Faculty of Science, Shizuoka University, Shizuoka, Japan. [7]Department of Biomolecular Science, Toho University, Funabashi, Japan. [8]Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan. [9]Department of Nephrology & Hypertension, Mayo Clinic College of Medicine & Science, Rochester, MN, USA. ✉e-mail: akirkato@bio.titech.ac.jp

interesting example to elucidate the cause of the secondary loss of an organ; nevertheless, the physiological benefits and developmental mechanisms involved in this secondary loss have not yet been clarified.

Genome sequences of many ray-finned fishes have been recently published and the number of species available allows some comprehensive analysis on the genomic difference between gastric and agastric fishes. In particular, agastric fishes from the following four orders of teleosts have been sequenced: zebrafish (*Danio rerio*; Cypriniformes)[17], Japanese medaka (*Oryzias latipes*; Beloniformes)[18], pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis*; Tetraodontiformes)[19,20], and wrasse (*Labrus bergylta*)[21]. Genome sequences of gastric fishes such as three-spined stickleback (*Gasterosteus aculeatus*)[22], Atlantic cod (*Gadus morhua*)[23], and Nile tilapia (*Oreochromis niloticus*)[24] have also been published. Based on these analyses, the H^+/K^+-ATPase (*atp4a* and *atp4b*) and pepsinogens (*pga*, *pgc*) genes are co-deleted in the genomes of agastric species but are present in the genomes of gastric species[4,21]. In monotremes, convergent gene losses for *atp4a*, *atp4b*, *pgc*, and *pga* occurred in platypus, and those for *pgc* and *pga* occurred in echidna, suggesting that the loss of *pgc* and *pga* occurred before the platypus-echidna split at more than 21 mya[16,25].

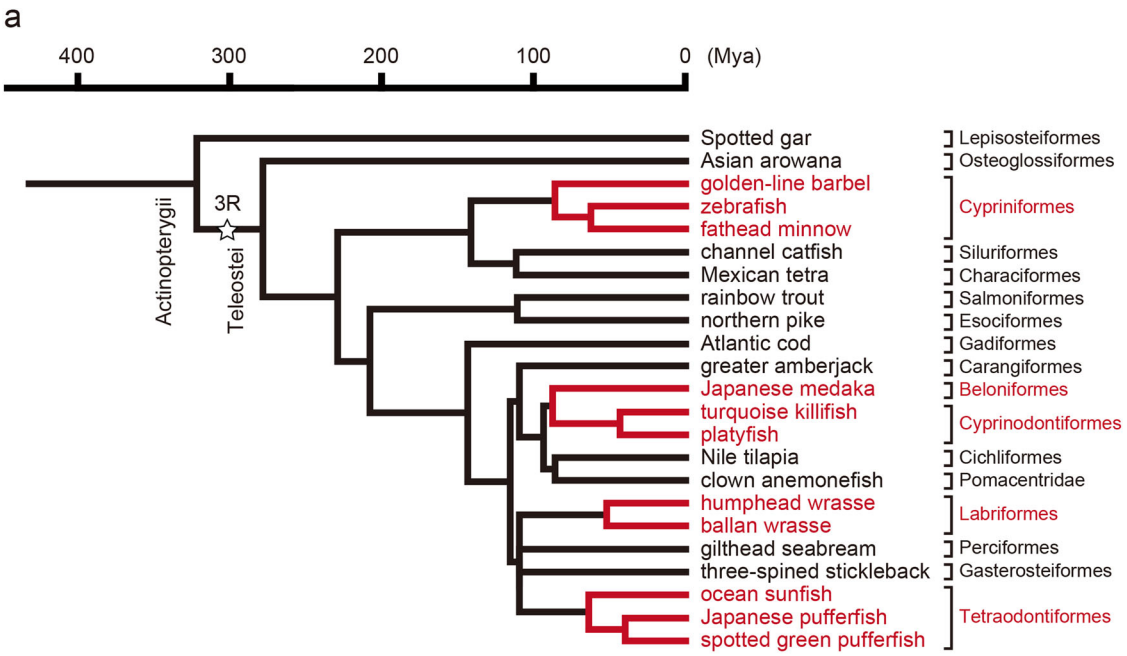
During our studies of anion transporters of solute carrier family 26 (Slc26) in pufferfish and eels[26,27], we found that the gene or cDNA for Slc26a9 was absent in the expressed sequencing tag (EST) and genome databases of pufferfish, zebrafish, and Japanese medaka, but present in those of three-spined stickleback, rainbow trout, Atlantic cod, and Nile tilapia. In mice, Slc26a9 is highly expressed in the stomach and lung[28], and its deletion causes tubulovesicle loss in parietal cells, acid[29] and prostaglandin-stimulated HCO_3^- secretion impairment in the stomach[30], and airway mucus obstruction through airway inflammation[31]. These results indicate that the absence of *slc26a9* in fish species is correlated with stomach loss, and that more genes that are important for gastric function could be lost among agastric fishes in a convergent manner. To confirm this hypothesis, we compared gene losses between agastric and gastric fishes and identified additional genes that are co-deleted in agastric fishes to demonstrate a novel genotypic convergence in relation to stomach loss.

Results

Screening of genes co-deleted in the genomes of agastric fishes

Genes which are commonly absent stomachless fish genomes were screened by database mining. First, a list of all annotated genes in the three-spined stickleback genome database[22] was obtained using the Ensembl BioMart tool[32] and compared to those of agastric fishes (zebrafish;[17], Japanese medaka;[18], spotted green pufferfish;[20], and Japanese pufferfish;[19]); approximately 80 three-spined stickleback genes were identified that were absent in the gene annotations of the agastric fishes. Second, the presence or absence of the identified genes was confirmed by a homology search in the genome databases for agastric fishes (zebrafish, Japanese medaka, spotted green pufferfish, and Japanese pufferfish). Blast analyses showed that many of those genes were present in agastric fishes but not correctly annotated or annotated with a different name. Ten genes, *atp4a*, *atp4b*, *pgc*, *slc26a9*, *kcne2*, *vsig1*, *pqlc2l*, *pradc1*, *atp6v0d2*, and *ankub1*, were confirmed to be absent in the genome of these agastric fishes but present in three-spined stickleback. A similar analysis was performed on 23 Actinopterygii species. Phylogenetic relationships among the 23 species are shown in Fig. 1a. Finally, six genes, *atp4a*, *atp4b*, *pgc*, *slc26a9*, *kcne2*, and *vsig1* were confirmed to be absent in the genome of the majority of the agastric fishes but present in gastric species. Three of these six genes (*atp4a*, *atp4b*, and *pgc*) were also

a



b

Gene	Product	Function
slc26a9	solute carrier family 26, member 9	anion transporter/channel
kcne2	potassium channel, voltage gated subfamily E regulatory beta subunit 2	potassium channel
cldn18a	claudin-18	tight junction protein
vsig1	V-set and immunoglobulin domain containing 1	membrane protein

Fig. 1 | Evolutional relationships of Actinopterygii gastric and agastric species and list of genes co-deleted in the genomes of agastric fishes identified in this study. a The time-calibrated phylogeny of 23 species analyzed in this study was prepared based on Near et al. [1,130] and the TimeTree database (http://www.timetree.org/)[131]. Species in lineages of the four agastric lineages (Cypriniformes, Beloniformes & Cyprinodontiformes, Tetraodontiformes, and Labriformes) are shown in brown. 3R, teleost-specific third-round whole genome duplication. **b** List of genes co-deleted in the genomes of agastric fishes identified in this study[29,48,49,132].

reported absent in agastric fishes by Castro et al.[4], corroborating the validity of this strategy. However, *pga2* was not included in the list, indicating the incompleteness of this method.

We next individually analyzed the presence of genes whose function or expression in the stomach of mammals was recognized using blast analyses. As previously reported[4], *pga2* was confirmed to be absent in the genomes of agastric fishes. In addition, a teleost fish-specific ohnolog of the claudin 18 gene, *cldn18a*, was found to be co-deleted in the genome databases of these fishes. Another ohnolog, *cldn18b* was shown to be present in gastric fishes and some agastric fishes (zebrafish and Japanese pufferfish). In total, four genes (*slc26a9*, *kcne2*, *cldn18a*, and *vsig1*) were found to be co-deleted in the genome databases of the most agastric fish species of Actinopterygii (Fig. 1b).

Identification of genes co-deleted in the genomes of agastric fishes

Synteny and dot plot analyses were performed to evaluate gene loss and pseudogenization, respectively. A synteny analysis of the four identified genes (*slc26a9*, *kcne2*, *cldn18a*, and *vsig1*) and the related *cldn18b* ohnolog was performed on 23 Actinopterygii species and shown in Fig. 2 and Supplementary Tables 1–5. The results of dot plot analyses are shown in Supplementary Figs. 1–4. A summary of the presence or deletion of each exon-coding region is shown in Fig. 3. The results showed that *kcne2*, *vsig1*, and *cldn18a* were absent or pseudogenized in all 11 species in four agastric lineages (Cypriniformes, Beloniformes and Cyprinodontiformes, Tetraodontiformes, and Labriformes) but present in all other species in 12 gastric lineages (Figs. 2b, c, e, and 3b–d). *slc26a9* was absent or pseudogenized in nine agastric species in three lineages (Cypriniformes, Beloniformes, Cyprinodontiformes, and Tetraodontiformes) but present in the other species including two species of Labriformes (wrasses) (Figs. 2a and 3a). The *cldn18b* ohnolog was deleted in seven species in two lineages (Beloniformes, Cyprinodontiformes, and Tetraodontiformes) but existed in the other species including four species in agastric lineages (Cypriniformes and Tetraodontiformes) (Fig. 2d).

Synteny and dot plot analyses of *atp4a*, *atp4b*, *pgc*, and *pga2* was similarly performed (Figs. 3e–h and 4, Supplementary Figs. 5–8, Supplementary Tables 6–10). The results revealed that *atp4a*, *atp4b*, *pga2*, and *pgc* were absent or pseudogenized in all 11 species in four agastric lineages but present in all the other species in 12 gastric lineages.

pga orthologs are distributed in three loci in the teleost genome and are named *pga1*, *pga2*, and *pga3*[4,33] (Fig. 4d, e). Phylogenetic analysis of *pga* orthologs and the details of the evolutionary relationships are shown in Fig. 5 and described in the next chapter. In contrast to *pga2*, which was deleted in all 11 agastric species, *pga1* was deleted in eight species of three agastric lineages (Cypriniformes, Beloniformes, and Labriformes) but existed in other species including Tetraodontiformes (Fig. 4d). The *pga3* gene was deleted in several agastric and gastric fishes[4] (Fig. 4e).

Evolution of *pga* in bony vertebrates

Although teleost fishes have three *pga* paralogs, *pga1*, *pga2*, and *pga3*[4,33], no clear evolutionary relationship among the paralogs has been uncovered. Therefore, a comprehensive analysis of *pga* was conducted on the genomic data of cartilaginous fish, tetrapods, lobe-finned fish, and ray-finned fish. The *pga* paralog nomenclature in representative species is shown (Fig. 5a), and a molecular phylogenetic tree was constructed (Fig. 5b). These results suggest that cartilaginous fish, tetrapods, lobe-finned fish, and ray-finned fish each have their own *pga* paralogs. In cartilaginous fishes, *pga* paralogs consist of four major branches, indicating that the divergence of these four branches occurred before the speciation of cartilaginous fishes, and that they acquired species-specific paralogs after speciation. The tetrapod *cym* is positioned as a tetrapod-specific paralog. The *pga* paralogs of the coelacanth, a lobe-finned fish, formed a single branch, suggesting that *pga* paralogs evolved independently in lobe-finned fish (Fig. 5b, clear highlight).

The *pga* paralogs of ray-finned fish formed five major branches, each of which contained *pga* from diverse species, suggesting that these five branches arose from the common ancestor of ray-finned fish. In gray bichir and reedfish, all *pga* paralogs were located in tandem (Fig. 5a), suggesting that these ancestral paralogs arose by tandem duplication. In this study, these ancestral *pga* paralogs were provisionally named *pga.r1*, *pga.r2*, *pga.r3*, *pga.r4* and *pga.r5*, with r1-r5 representing paralogs arising from ray-finned fish-specific tandem duplications. Synteny of extant *pga* derived from *pga.r1*-*pga.r5* is shown in Fig. 5a. Gray bichir, for example, has one ortholog derived from *pga.r1*, *pga.r2*, *pga.r4*, and *pga.r5*, and four from *pga.r3*. The spotted gar had one ortholog derived from *pga.r1* and three from *pga.r3*. All previously named *pga1*, *pga2*, and *pga3* in teleost fish are orthologs derived from *pga.r3*. Many teleosts only have orthologs derived from *pga.r3*, whereas the European eel has orthologs derived from *pga.r1* and *pga.r3* and the Indo-Pacific tarpon has orthologs derived from *pga.r1*, *pga.r3*, and *pga.r4*. These results can be considered an example of a birth-and-death model in gene family evolution[34]. In the phylogenetic tree, we included the amino acid sequence derived from the *pga2* pseudogene (*pga2-ps*) of ocean sunfish. Ocean sunfish *pga2-ps* was positioned in the teleost *pga2* group with a long branch.

Expression of stickleback genes whose orthologs are deleted in agastric fishes

Various three-spined stickleback tissues were analyzed by semi-quantitative RT-PCR to determine the distributions of mRNAs for the eight genes (Fig. 6a), as well as for *actb* as a positive control showing cDNA integrity. The results showed that *atp4a*, *atp4b*, *kcne2*, *slc26a9*, *vsig1*, *cldn18a*, *pgc*, and *pga2* were highly expressed in the stomach. Several of these genes were also expressed in stickleback organs other than the stomach: *kcne2* was observed in the ovary and testis, *pgc* in the gut and liver, *pga* in various organs, including the gut, liver, and kidney, *vsig1* in the gut and liver, and *cldn18a* in the gut.

To identify the cells expressing the genes at the tissue level, in situ hybridization and histology were performed on the three-spined stickleback gut (Fig. 7), which is composed of a mucosa, submucosa, muscularis, and serosa (Fig. 7a, b). The mucosa consists of a gastric pit and gastric (oxyntic) gland in the anterior cardiac or fundic region of the stomach, and a gastric pit only in the posterior pyloric region. All genes tested were expressed in the mucosa of the three-spined stickleback stomach, with none expressed in the other layers. All eight genes, *atp4a*, *atp4b*, *pgc*, *pga2*, *slc26a9*, *kcne2*, *cldn18a*, and *vsig1*, were expressed in gastric gland cells (Fig. 7c, d), and three *pga2*, *cldn18a*, and *vsig1*, were expressed in the columnar mucous cells of the gastric pit (Fig. 7c, e) which had characteristic Periodic acid-Schiff (PAS)-positive mucous granules in the apical region (Fig. 7b). Hybridization using sense probes did not resulted in any labeling (Supplementary Fig. 9). In general, the gastric gland of fishes consists of only one secretory cell type (oxynticopeptic cells), whereas that of mammals is composed of chief cells for digestive-enzyme secretion and parietal cells for acid secretion[5]. In the gastric gland of the three-spined stickleback, most epithelial cells presented positive expressions for genes involved in acid secretion (*atp4a*, *atp4b*, *slc26a9*, *kcne2*) and digestive enzymes (*pgc*, *pga2*) (Fig. 7c, d), indicating that these eight genes are coexpressed in three-spined stickleback oxynticopeptic cells.

Expression of wrasse *slc26a9*

Intact *slc26a9* was present in wrasses but not in the other agastric species (Figs. 2a and 3a). To confirm whether *slc26a9* is transcribed in organs other than the stomach, total RNA was extracted from various organs of a humphead wrasse and semi-quantitative RT-PCR was performed. In the humphead wrasse, *slc26a9* was expressed in the eyes, gills, fins, and skin (Fig. 6b).

Rapid evolution of *cldn18b* in agastric fishes

Gastric teleosts have two orthologs for claudin 18, *cldn18a* and *cldn18b*, whereas agastric teleosts have a single or deleted claudin 18 gene. The paralogs are specifically present in Teleostei but not in tetrapods. For both *cldn18a* and *cldn18b* loci, the synteny of the neighboring genes, *hs2st1*/*hs2st1a* and *sox14*, are conserved (Fig. 2c, d). These results indicate

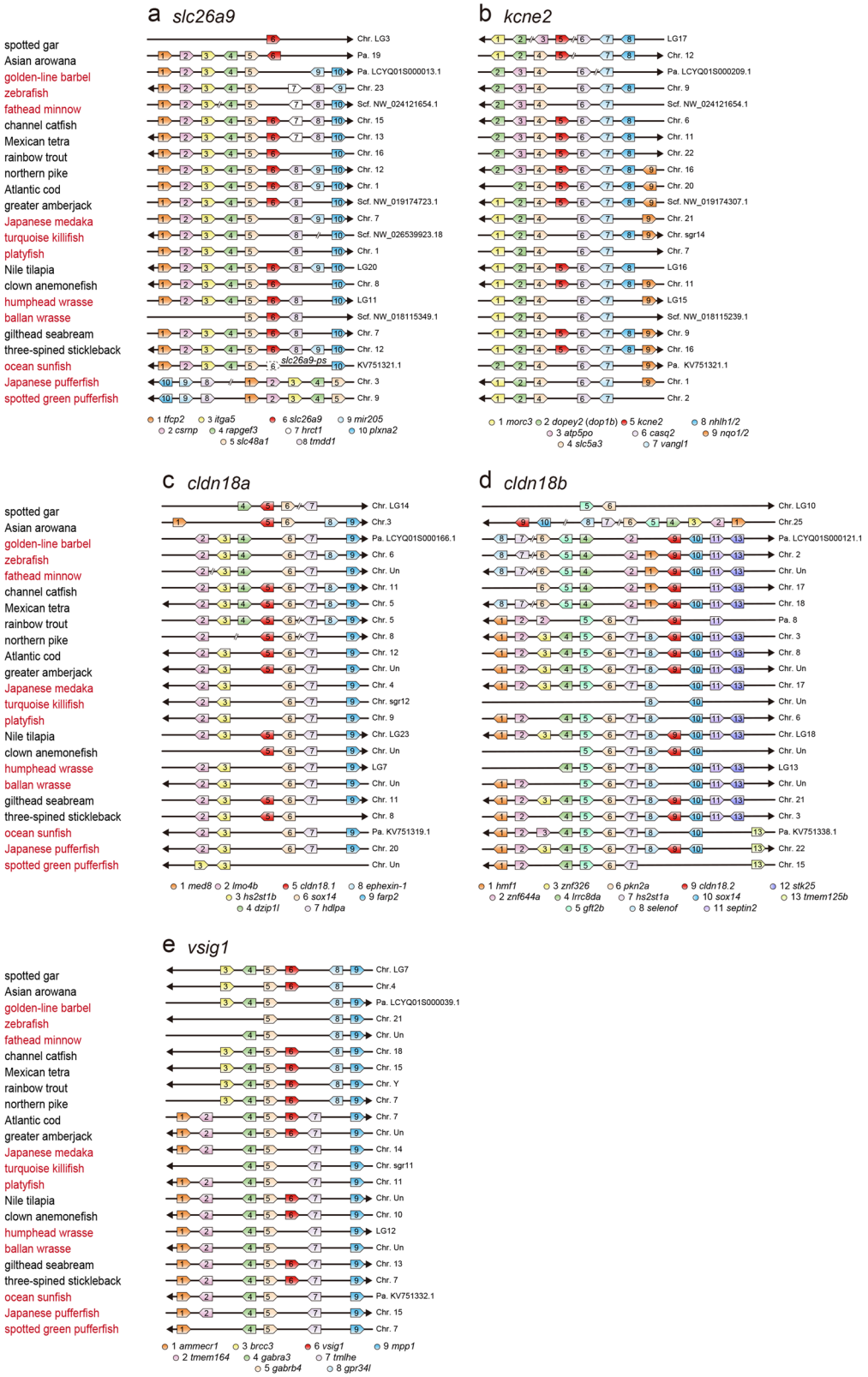


Fig. 2 | Co-deletion of genes in genomes of agastric fishes found in this study. The synteny analyses of *slc26a9* (**a**), *kcne2* (**b**), *cldn18a* (**c**), *cldn18b* (**d**), *vsig1* (**e**) in the genome databases of 23 Actinopterygii species are shown. Names of species of the four agastric lineages (Cypriniformes, Beloniformes and Cyprinodontiformes, Tetraodontiformes, and Labriformes) are shown in red. Chr. chromosome; Scf. scaffold. Accession numbers of each gene are shown in Supplementary Tables 1–5.

that *cldn18a* and *cldn18b* are ohnologs that are generated by teleost-specific genome duplication (TGD)[35]. The presence of *cldn18a* is highly associated with the existence of a stomach, whereas the presence of *cldn18b* is only partially associated with the possession of this organ.

To compare the evolution of *cldn18* between animals with and without a stomach, mean rates for non-synonymous and synonymous substitutions, d_n and d_s, respectively, were calculated for four groups: (i) *cldn18* of tetrapods/coelacanths, (ii) *cldn18a* of gastric fish, (iii) *cldn18b* of gastric fish, and

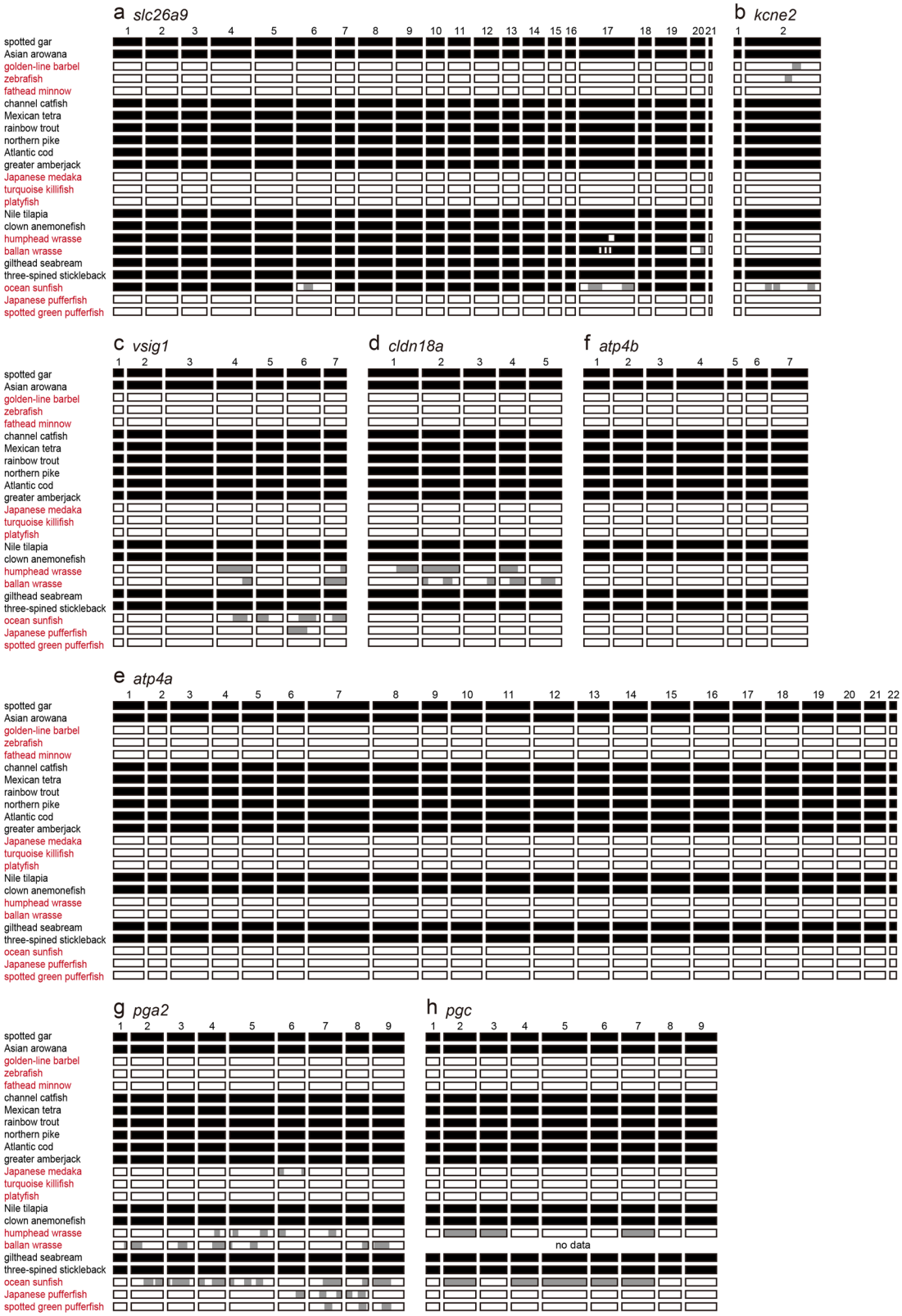


Fig. 3 | Genetic deletions or changes in agastric fishes. Deletions in the exons of *slc26a9* (**a**), *kcne2* (**b**), *vsig1* (**c**), *cldn18a* (**d**), *atp4a* (**e**), *atp4b* (**f**), *pga2* (**g**), and *pgc* (**h**) in 23 ray-finned fish species. Schematic representations of the dot plot analyses (Supplementary Figs. 1–8) are shown. The presence or absence of exons of each gene are indicated by black and white boxes, respectively. Partially homologous exons are shown by gray boxes. Species in the four agastric lineages are shown in red.

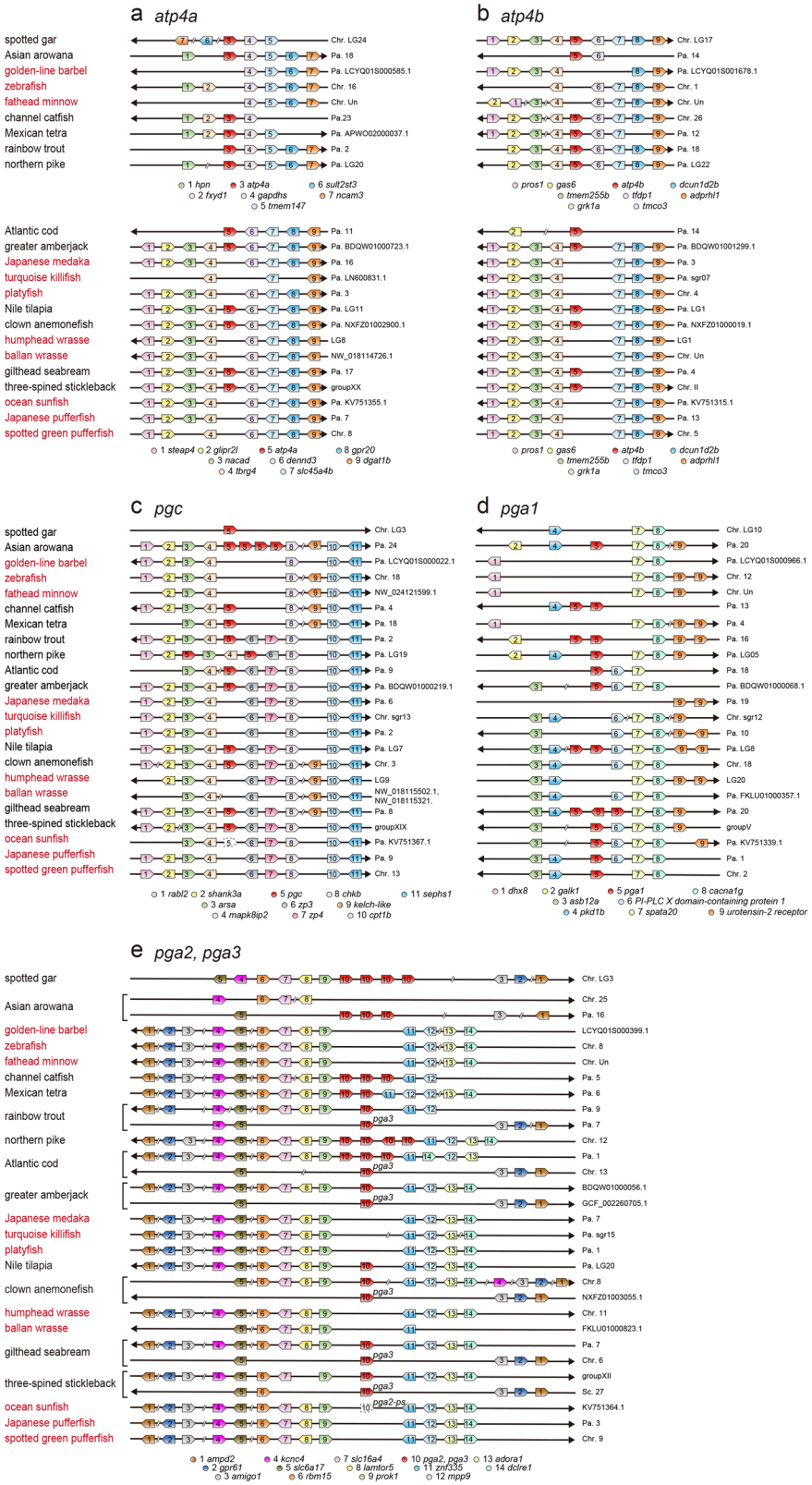


Fig. 4 | Synteny analyses of genes known to be co-deleted in the genomes of agastric fishes. The synteny analyses of *atp4a* (**a**), *atp4b* (**b**), *pgc* (**c**), and *pga* (**d**, **e**) in the genome databases of 23 Actinopterygii species are shown. Names of species of the four agastric lineages (Cypriniformes, Beloniformes and Cyprinodontiformes, Tetraodontiformes, and Labriformes) are shown in red. Chr., chromosome; Scf., scaffold. Accession numbers of each gene are shown in Supplementary Tables 6–10.

(iv) *cldn18b* of agastric fish (zebrafish and Japanese pufferfish). Non-synonymous substitutions occurred ~4 times more frequently in the *cldn18b* of agastric fishes than in the other groups ($P < 0.0001$; Fisher's exact probability test) (Fig. 6c–e). These results suggest that the loss of the stomach allows higher amino acid substitution rates on *cldn18b*, which is likely due to the relaxation of functional constraints.

Pseudogenization of *vsig1* in platypus and loss of *kcne2* in echidna

As reported previously[16,25], we confirmed convergent gene losses of *atp4a*, *pgc*, and *pga* in the platypus and echidna (Fig. 8a–d). In both organisms, *atp4b* was annotated in the genome database (XM_039915013.1, and XM_038761646.1, respectively); however, the predicted amino acid

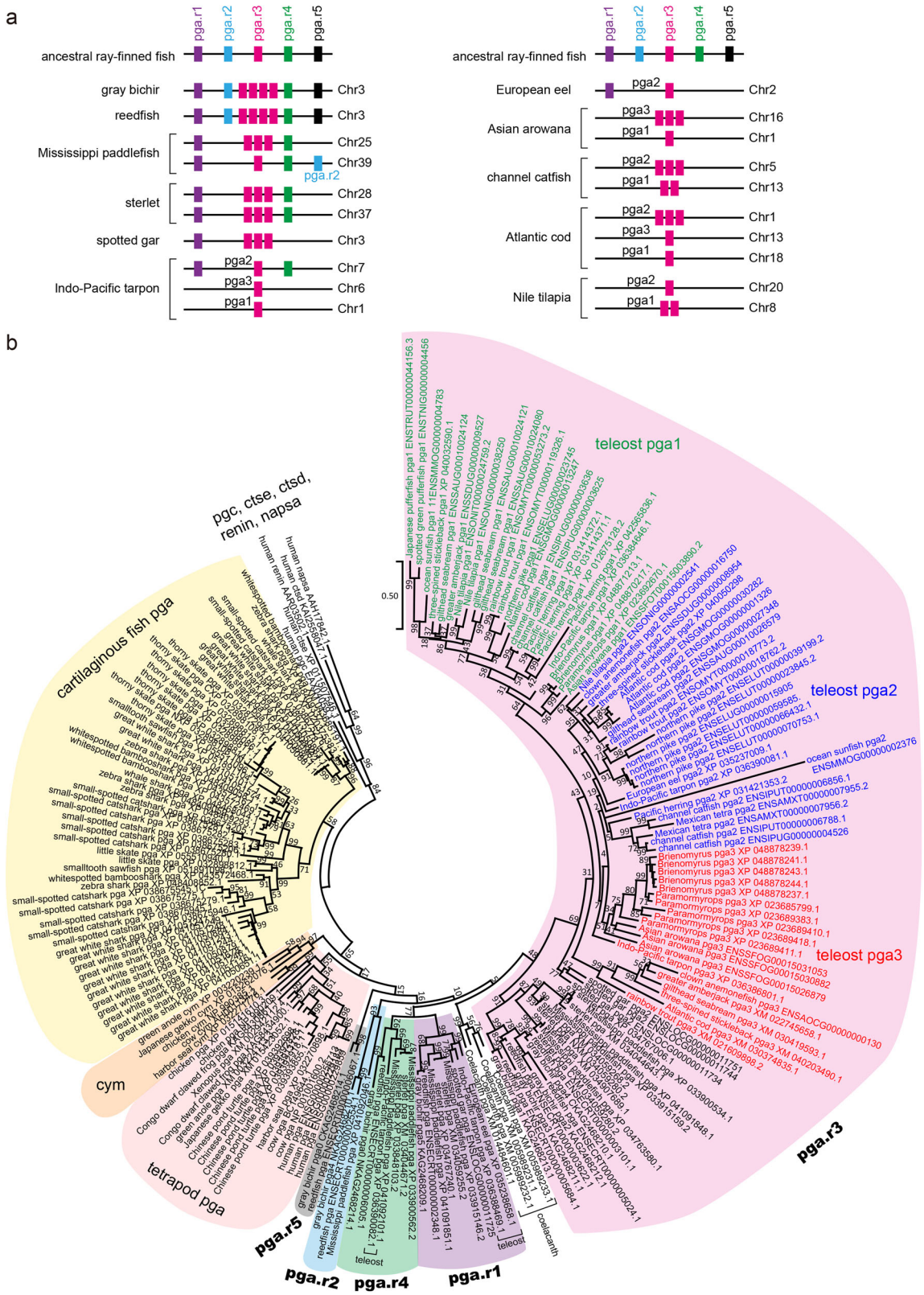


Fig. 5 | Evolution of *pga* orthologs in ray-finned fishes. a Schematic representation of the synteny of ray-finned fish-specific paralogs of *pga*. Putative ancestral paralogs in ray-finned fishes are indicated as *pga.r1*– *pga.r5*. Synteny of *pga* orthologs in representative species are shown. **b** Phylogenetic analysis of *pga* and related genes in ray-finned fishes, coelacanth, tetrapods, and cartilaginous fishes. The deduced amino acid sequence of each gene was aligned using ClustalW software and a phylogenetic tree was constructed using the maximum-likelihood method with MEGA software. Numbers indicate bootstrap values, and the scale bar represents the genetic distance of amino acid substitutions per site. A list of genome databases used for the analysis is shown in Table 1.

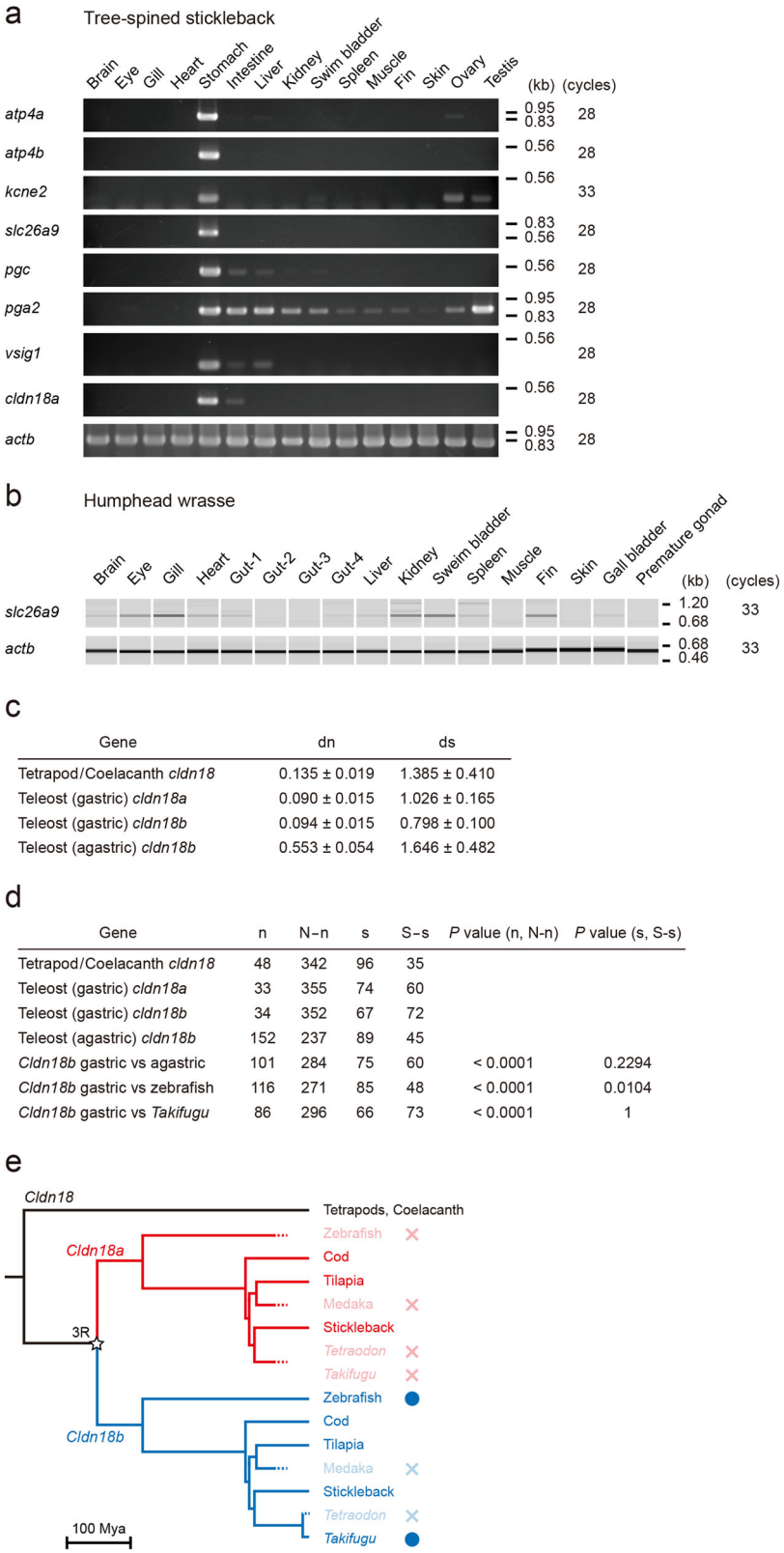


Fig. 6 | Expression of the three-spined stickleback and humphead wrasse genes whose orthologs are co-deleted in agastric fishes and rapid evolution of *cldn18b* in agastric fishes. a Expression of the three-spined stickleback genes whose orthologs are co-deleted in agastric fishes. Reverse transcription-PCR was performed on total RNAs purified from various three-spined stickleback tissues, and analyzed by agarose gel electrophoresis. **b** Expression of humphead wrasse *slc26a9*. Reverse transcription-PCR was performed on total RNAs purified from various humphead wrasse tissues, and the pseudo-gel images of PCR products were generated using a microchip electrophoresis system. *actb* was used as an internal control in each species. Numbers indicate the PCR cycles. **c** Average Jukes-Cantor (JC) distances of claudin 18 coding regions within or among the groups. Variants were estimated using the bootstrap method with 500 replicates. Nucleotide sequences from three species for tetrapod/coelacanth *cldn18*, three species for gastric teleost *cldn18a*, three species for gastric teleost *cldn18b*, and two species for agastric teleost *cldn18b* were used for the analysis. dn, non-synonymous substitutions per site; ds, synonymous substitutions per site. **d** Average numbers of non-synonymous differences (n), synonymous differences (s), unchanged non-synonymous sites (N-n), and unchanged synonymous sites (S-s) of claudin 18 coding regions within or among the groups. *P*-values from two-sided Fisher's exact test are shown. **e** Evolutionary model of *cldn18* ohnologs in Teleostei. Time-calibrated phylogeny was prepared based on the reports of Near et al.[1,130] and Kumar and Hedges[133]. Red and blue indicate the two ohnologs, *cldn18a* and *cldn18b*, respectively. Gene losses are indicated by dotted lines in the tree and x marks. Blue circles indicate rapidly evolved genes. An open star indicates teleost-specific third-round whole genome duplication[35].

sequences lacked the amino-terminal cytoplasmic and the transmembrane domains (Supplementary Fig. 10), which are encoded by the exons 1 and 2 of *atp4b* in other species. TBLASTN analysis of the whole-genome databases of platypus and echidna did not reveal regions encoding the cytoplasmic and transmembrane domains of Atp4b. Because Atp4b is a membrane protein with one transmembrane domain[36], *atp4b* is considered to have lost its original function in the platypus and echidna and may be pseudogenized in these species (Fig. 8b).

The presence or absence of the four genes (*slc26a9*, *kcne2*, *cldn18*, and *vsig1*) was searched using the genome databases of the coelacanth[37], *Xenopus*[38], anole lizard[39], platypus[40], echidna[25], and human[41] by blast searches of their genome sequences. All genes were present in the genomes of the

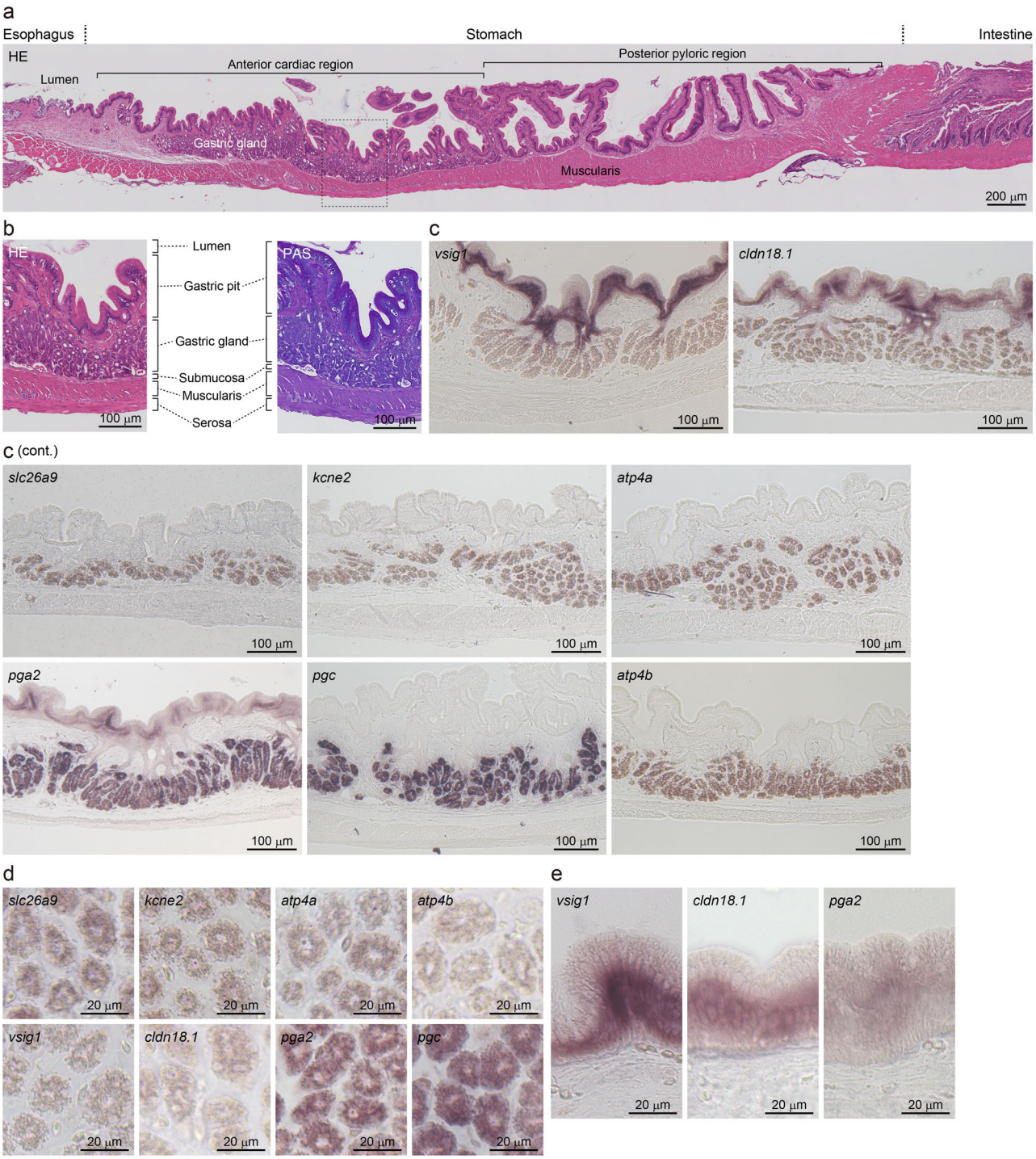
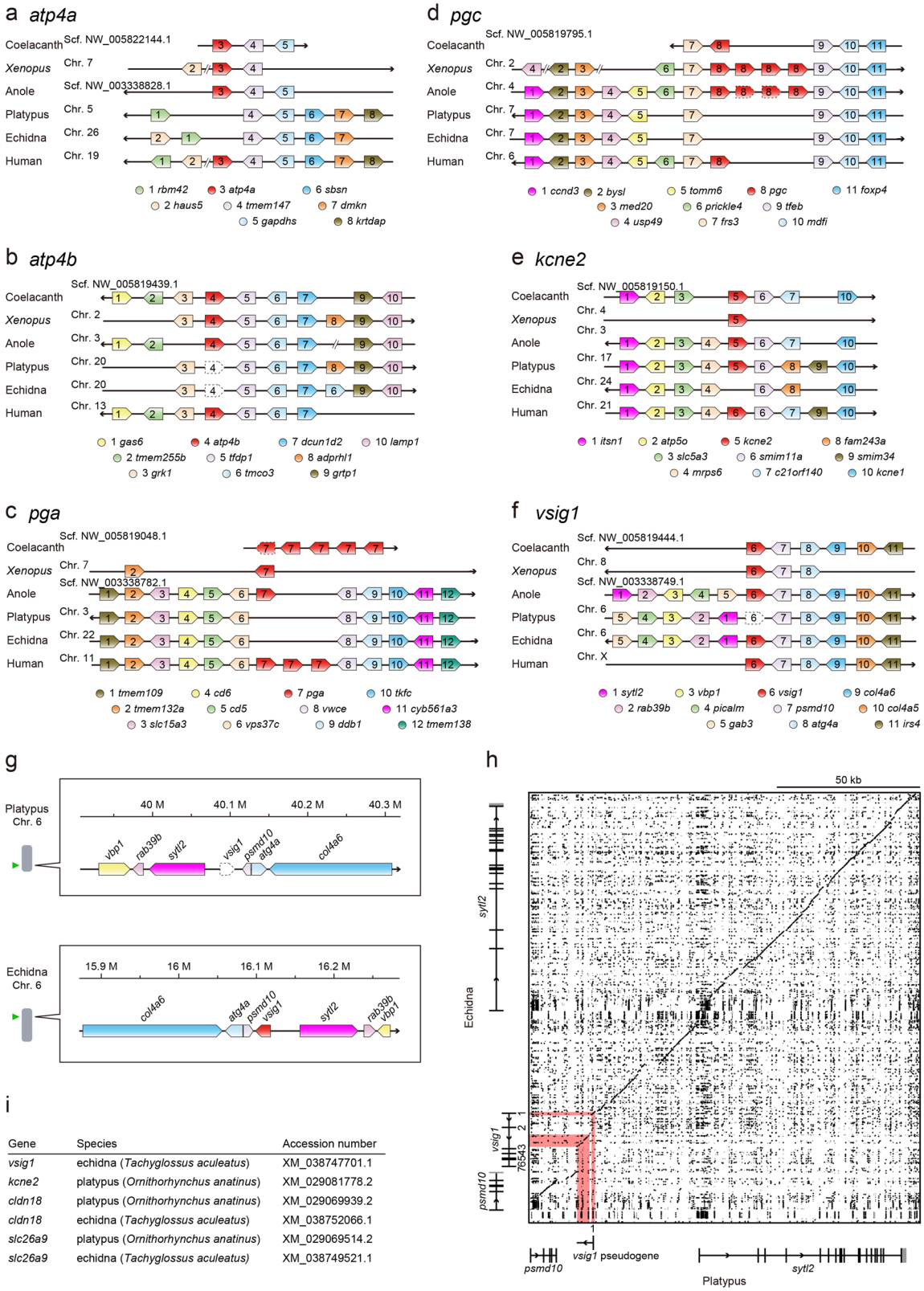


Fig. 7 | In situ hybridization histochemistry analysis of the three-spined stickleback stomach for the genes whose orthologs are co-deleted in gastric fishes. a A vertical section of the whole stomach stained with hematoxylin and eosin. **b** Large magnification views of gastric wall sections stained with hematoxylin and eosin (left) or Periodic acid-Schiff reagent (right). **c**–**e** In situ hybridization. The gastric wall sections were stained with antisense probes. Results with sense probes for the negative control are shown in Supplementary Fig. 1. Large magnification views for the gastric gland and gastric pit are shown in (**d**, **e**), respectively. HE Hematoxylin and eosin, PAS Periodic acid-Schiff.

gastric species, coelacanth, *Xenopus*, anole lizard, and human. *cldn18* and *slc26a9* were retained in the genomes of both platypus and echidna (Fig. 8i). Convergent gene loss for *kcne2* was observed in the echidna, but not in the platypus (Fig. 8e). *vsig1* was pseudogenized in the platypus but not in the echidna (Fig. 8f), and dot plot analysis showed a pattern of deletion of *vsig1* in the platypus, with exons 2–7 deleted at the homologous locus of *vsig1* (Fig. 8g–h).

Discussion

Genome projects of vertebrate species have allowed the clarification of the presence of lineage-specific gene losses during evolution[42–47]. In the present comparative genomic analysis, the deletion of four genes was shown to be associated with secondary stomach losses in Actinopterygii species. The four genes contain the Cl^- channel-transporter (*slc26a9*) and a regulatory subunit of the K^+ channel (*kcne2*). These molecules are co-expressed with



Gene	Species	Accession number
vsig1	echidna (*Tachyglossus aculeatus*)	XM_038747701.1
kcne2	platypus (*Ornithorhynchus anatinus*)	XM_029081778.2
cldn18	platypus (*Ornithorhynchus anatinus*)	XM_029069939.2
cldn18	echidna (*Tachyglossus aculeatus*)	XM_038752066.1
slc26a9	platypus (*Ornithorhynchus anatinus*)	XM_029069514.2
slc26a9	echidna (*Tachyglossus aculeatus*)	XM_038749521.1

Fig. 8 | Loss of genes in the genomes of monotremes. The synteny analyses are shown. **a–d** Loss or pseudogenization of *atp4a*, *atp4b*, *pgc*, and *pga2* in platypus and echidna[16,25]. **e** Loss of *kcne2* in echidna but not in platypus. **f** Pseudogenization of *vsig1* in platypus but not in echidna. **a–f** Arrowheads represent the right and left orientations, respectively, of the genome sequences in the NCBI and ENSEMBLE databases. Arrow-shaped boxes indicate the orientation of each gene. Arrow-shaped dotted box indicates pseudogene. Chr., chromosome; Scf., scaffold, Ctg, contig. Accession numbers of each gene are shown in Supplementary Tables 11–16. **g** Chromosomal localization of echidna *vsig1* and platypus *vsig1* pseudogene. **h** Dot plot analysis of the echidna *vsig1* and their flanking regions in comparison with the corresponding genome regions of the platypus containing a *vsig1* pseudogene. Homologous regions were plotted with dotmatcher program (window size: 20; threshold: 70). **i** *vsig1*, *kcne2*, *cldn18*, and *slc26a9* in monotremes.

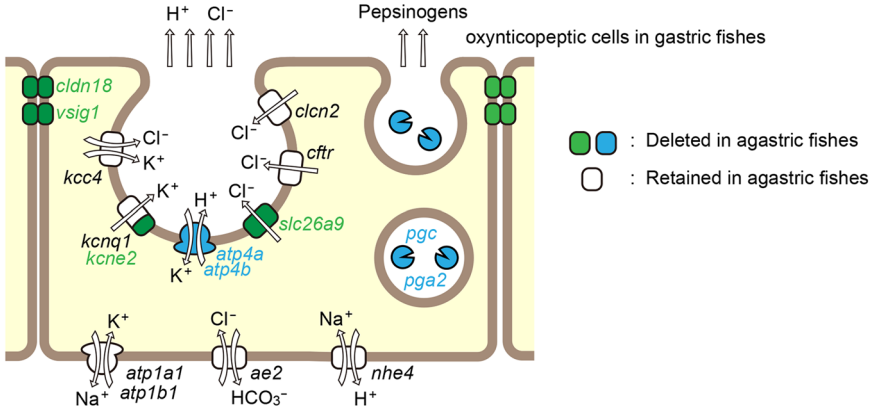


Fig. 9 | Schematic representation of functions of gastric proteins whose genes are deleted in agastric fishes. Epithelial model for the secretion of gastric acid and digestive enzymes in a fish oxynticopeptic cell is shown. Proteins whose genes were found to be lost in agastric fishes in the previous[4] or present study are indicated by blue and green, respectively. Gastric proteins that are retained in agastric fishes are illustrated in white.

H^+/K^+-ATPase in gastric gland cells of the stomach and are involved in gastric acid (HCl) secretion. The four genes also contain cell-cell adhesion molecules that are involved in the paracellular barrier function against H^+ (*cldn18*)[48] and control the stomach development (*vsig1*)[49]. These results, along with those of other studies on the deletion of genes for H^+/K^+-ATPase (*atp4a* and *atp4b*) and pepsinogens (*pga*, *pgc*)[4], we summarized the convergent losses of important functional genes in four major independent groups of agastric fishes, Cypriniformes (golden-line barbel, zebrafish, and fathead minnow), Beloniformes and Cyprinodontiformes (Japanese medaka, turquoise killifish, and platyfish), Tetraodontiformes (ocean sunfish, Japanese pufferfish and spotted green pufferfish), and Labriformes (humphead wrasse and ballan wrasse). *slc26a9* was present in wrasses and was expressed in organs other than the stomach, such as the gills and skin (Fig. 6b). This result suggests that an unidentified non-gastric function of *slc26a9* prevents its loss from wrasses.

Ocean sunfish (*Mola mola*) belongs to Tetraodontiformes and is closely related to pufferfishes. There is no histological analysis that clarify the presence or absence of gastric glands in the gut of ocean sunfish. In the digestive tract of ocean sunfish, a stomach-like organ is present[50]. However, the present analysis indicates that the genome of ocean sunfish has a similar pattern of gastric gene deletions as pufferfishes and other agastric fishes. This result suggests that the ocean sunfish may be an agastric fish. A stomach-like organ is also present in pufferfishes and is known as the abdominal pouch[51]. The abdominal pouch of pufferfishes is often called stomach and can temporarily store food, but the abdominal pouch does not have gastric glands nor the ability to digest food. In the case for ocean sunfish, further analysis is required to clarify the presence or absence of gastric glands in the stomach-like organ.

Gastric H^+ secretion is mediated by apical (luminal) H^+/K^+-ATPase coupled with the K^+ channel/transporter for K^+ recycling and is also accompanied by Cl^- secretion mediated by the apical Cl^- channel/transporter. In mammals, the Cftr, Clc-2, and Slc26a9 Cl^- channels are proposed to mediate Cl^- secretion[52,53]. K^+ is recycled by a K^+ channel composed of Kcnq1 α and Kcne2 β subunits. In addition, the apical K^+-Cl^- cotransporter (Kcc4) secretes K^+ and Cl^- together. Among the apical components for gastric acid secretion, four genes, *atp4a*, *atp4b*, *slc26a9*, and *kcne2* are deleted in agastric fishes, suggesting that the function of those genes is closely associated with gastric acid (H^+) secretion. The remaining genes were retained in agastric fishes, suggesting that they have important functions in non-gastric tissues of the agastric fishes. In non-gastric tissues, Cftr excretes Cl^- in the gills of marine teleosts and secretes intestinal Cl^-[54–56]. Kcc4 is involved in H^+ secretion in the renal α-intercalated cells in mammals[57], which may explain why these genes are retained. The lost genes code for some of the apical components but not the basolateral components such as Na^+/K^+-ATPase, anion exchanger 2 (Ae2), and Na^+/H^+ exchanger 4 (Nhe4) for gastric acid secretion (Fig. 9). In general, the basolateral membrane of epithelial cells faces the extracellular fluid with a stable ionic composition, whereas the apical membrane faces the luminal fluid with a variable composition. Therefore, functional proteins on the apical membrane tend to be tissue-specific, while those on basolateral membrane are shared among epithelia of various tissues. Our results suggest that the basolateral components for gastric acid secretion are common with those of other epithelial systems, thereby preventing the deletion of these genes, whereas some apical components are specific to the stomach, which are more prone to gene losses.

Our analysis revealed that the platypus genome contains *kcne2*, *slc26a9*, and *cldn18*, whereas the echidna genome contains *vsig1*, *slc26a9*, and *cldn18*. In mammals, *kcne2*[58,59], *slc26a9*[28,29,60,61], and *cldn18*[62,63] are expressed in the lung at high levels as well as in the stomach and their functions are related to both gastric and pulmonary systems. Slc26a9 is critical for respiratory function in terrestrial vertebrates as loss of *slc26a9* can create a cystic fibrosis-like phenotype[64–66]. In contrast in the three-spined stickleback, these genes are expressed in the stomach but not in the swim bladder or gill, which are related to respiratory function. These results suggest that *kcne2*, *slc26a9*, and *cldn18* are required mainly for gastric function in Actinopterygii, with the exception of wrasse slc26a9, which has non-gastric functions, whereas those are required for the gastric and pulmonary functions both in terrestrial vertebrates. Therefore, in platypus, the respiratory function of *slc26a9*, *kcne2*, and *cldn18* in the lung may prevent the loss of these genes. In echidnas, the respiratory function of *slc26a9* and *cldn18* in the lung may also prevent the loss of these genes. However, *kcne2* was lost in the echidna, suggesting that the respiratory function of *kcne2* was compensated for by another gene in this organism. RT-PCR analysis of sticklebacks showed that *kcne2*, *pga*, *pgc*, *pga*, *vsig1*, and *cldn18a* were expressed not only in the stomach but also in other organs. This result suggests that these genes function in organs other than the stomach of fish. However, in most agastric fish, these genes were deleted, probably because these functions were compensated for by another gene.

The loss of *vsig1* was observed in agastric fishes and platypus, but not in echidna. Vsig1 is a cell surface protein characterized by two extracellular immunoglobulin-like domains whose physiological function is still largely unknown. Vsig1 is also known as glycoprotein A34 (Gpa34) of tumor cells[67], is expressed in low- or non-metastatic cancer cells[68], and inhibits Yap/Taz signaling. Yap and Taz are transcriptional regulators and essential for cancer initiation or growth of most solid tumors[69]. As the Yap/Taz signaling is important for organogenesis[70], the role of Vsig1 for normal stomach development could be via the TAP/TAZ signaling[49]. In human and mice, the *vsig1* gene is expressed in the stomach and testes[49,67], while it was expressed in the stomach, intestine, and liver, but not in testes or other organs in the three-spined stickleback (Fig. 6a). Our result also indicated that Vsig1 is localized at the gastric gland and pit cells, which is identical to the case in mice[49]. The loss of *vsig1* could impair the development of stomach in platypus. Although the *vsig1* is an intact gene in echidna, their stomach is glandless. In this case, Vsig1 could be involved in the development of the stomach but some other factors control the development of the gastric gland.

Retention of *cldn18b*, a duplicated *cldn18* in teleosts, by some agastric fishes is a good example of how a gene evolves when the functional constraint is reduced. In the gastric epithelium, paracellular H^+ leakage is prevented by the tight junctions and associated junctional complexes, e.g., claudins. Only one component, claudin-18, has been identified as the paracellular H^+ barrier[48]. Complete deletion of both *cldn18a* and *cldn18b* in the genomes of Japanese medaka, turquoise killifish, platyfish, wrasses, ocean sunfish, and spotted green pufferfish indicates that the secondary stomach loss reduced the functional constraint of the *cldn18* genes. The *cldn18b* that is retained in some other agastric species (golden-line barbel, zebrafish, fathead minnow, and Japanese pufferfish) exhibited rapid non-synonymous substitution rates, which were higher than those of gastric species. Although *cldn18b* is retained in Japanese pufferfish, no tissues expressed the gene[71]. However, in zebrafish, *cldn18b* is also expressed in the kidney[72]. In mouse kidney, *cldn18* is expressed in the thick ascending limb of Henle's loop (TAL), which additionally expresses *cldn10*, *cldn16*, and *cldn19*. The mouse TAL functions as a site for the reabsorption of Ca^{2+} and Mg^{2+} via the paracellular pathway. In the mouse TAL, claudin-10 (claudin-10a: anion permeability; claudin-10b: cation ($Na^+ > K^+$) permeability) and -18 may contribute to the maintenance of barrier function, and claudin-16 and -19 contribute to Ca^{2+} and Mg^{2+} ion selectivity[73–75]. Because zebrafish kidneys also expresses claudin-10b[72], zebrafish claudin-18, together with claudin-10b and others, may contribute to the maintenance of tubular barrier function.

Many vertebrates have multiple *pga* gene paralogs. It is difficult to evaluate the evolutionary relationships of paralogs using the names of genes, as they are a mixture of those arising from old and new gene duplications. Castro et al. named *pga1*, *pga2*, and *pga3* as *pga* paralogs in three loci of the teleost genome[4]. Molecular phylogenetic analyses involving *pga* genes in cartilaginous fish, tetrapods, lobe-finned fish, and ray-finned fish have shown that teleost *pga1*, *pga2*, and *pga3* differ from *pga* paralogs in ancient ray-finned fishes, such as Polypterus, sturgeon, and gar. This confirmed that *pga1*, *pga2*, and *pga3* are teleost-specific paralogs, as reported by Castro et al.[4]. Interestingly, of the four *pga* paralogs in spotted gar (provisionally named *Locpga1*, *Locpga2*, *Locpga3*, and *Locpga4*), *Locpga1* belonged to the same branch as the *pga* paralogs of polypterus and sturgeons; *Locpga2*, *Locpga3*, and *Locpga4* belonged to the same branch as teleosts *pga1*, *pga2*, and *pga3*, which are paralogs that arose after the divergence of gar and teleosts. Synteny analysis suggested that *pga2* and *pga3* are present in loci generated by teleost-specific genome duplication (TGD); however, it remains unclear whether *pga2* and *pga3* are ohnologs or paralogs derived from pre-TGD tandem duplication. Species- and lineage-specific tandem duplications of *pga2* have been observed in various species (e.g., channel catfish, Mexican tetra, northern pike, and Atlantic cod). In the present analysis, *pga2* was the *pga* family member whose absence was most frequently associated with secondary loss of the stomach, whereas *pga1* and *pga3* were also observed in various gastric fishes. *pga1* synteny was conserved in many teleost species, although no synteny was observed with teleost *pga2*, *pga3*, or tetrapod *pga*. Given this, and the fact that *pga1* is a teleost-specific paralog, it is possible that *pga1* arose in the common ancestor of teleost fish via duplication through translocation. Among teleost fishes, *pga1* was present in most gastric fishes and some agastric fishes and was absent in some gastric fishes and many agastric fishes. In the agastric Japanese pufferfish, *pga1* is expressed in non-gastric tissues such as the skin[76].

The physiological advantages of secondary stomach loss are still largely unknown[5]. In the treatment of human gastric cancer, gastrectomy alters physiological properties such as oxygen availability, pH, food transit time, intestinal motility, and hormonal conditions, and alters the overall microbiome community structure[77]. Gastrectomy-associated alterations in microbial functions, such as nutrient transport and biosynthesis of organic compounds, may be related to changes in post-gastrectomy metabolism. In gastric teleost species, the stomach has a variety of physiological functions, such as food digestion, temporal food storage, pathogen invasion defense, and hormonal secretion[5]. The differences in the physiological properties between gastric and agastric fish remain unclear. As the stomach kills microorganisms using gastric acid and provides increased uniformity in the population of gut microbes[78], it is presumed that loss of the secondary stomach has some effect on the gut microbiome, and that the gut microbiome of agastric fish is more susceptible to environmental influences. Studies on the fish digestive tract microbiome indicate that fish harbor specialized gastrointestinal microbial communities like other vertebrates such as mammals[79–81], and the gut microbiomes of wood-eating catfishes, zebrafish, guppies, and others are related to their diets[79,82–85]. Further studies are required to better understand the physiological advantages of losing the secondary stomach.

Our results raise the question of whether the gene deletions observed in this study caused the stomach loss, or whether the deletions occurred after the stomach loss. Despite stomach loss, our study did not show deletion of the genes for transcriptional or growth factors that regulate stomach development in agastric fishes[86–88]. Thus, it is conceivable that the lack of a stomach is associated with the malfunction of the cis-regulatory elements for stomach development, which cannot be identified using the current strategy. It is also possible that a deletion of one of the eight genes caused a depletion of stomach function in fishes for which this depletion was neutral or advantageous, and additional gene deletion followed, causing the stomach to be completely regressed in the gut of fishes.

In conclusion, we identified novel genes that were lost in agastric fishes among four major teleost lineages, which suggests a convergent evolution scenario in relation to stomach loss. These genes encode apical ion channels for gastric acid secretion, and the cell-cell adhesion molecule that forms the paracellular H^+ barrier in the gastric epithelium (Fig. 9). These results indicate that a common cassette of gene losses occurred independently during or after stomach loss in the several agastric fish groups. Further studies are required to identify the causative genotype that triggered this stomach loss.

Methods

Screening of genes co-deleted in the genomes of agastric fishes

Lists of all annotated genes in the genome databases for zebrafish (*Danio rerio*)[17], Atlantic cod (*Gadus morhua*)[23], Nile tilapia (*Oreochromis niloticus*)[24], Japanese medaka (*Oryzias latipes*)[18], three-spined stickleback (*Gasterosteus aculeatus*)[22], Japanese pufferfish (*Takifugu rubripes*)[19], and spotted green pufferfish (*Tetraodon nigroviridis*)[20] were downloaded from Ensembl (http://www.ensembl.org/index.html)[89] using Ensembl BioMart tool[32]. After removing characters that indicated gene duplications, the presence or absence of all annotated three-spined stickleback genes in agastric fishes (zebrafish, Japanese medaka, spotted green pufferfish, and Japanese pufferfish) were determined through a text search using Excel software (Microsoft, Redmond, WA, USA). From this data, a list of three-spined stickleback genes that were commonly lacking in the gene lists of zebrafish, Japanese medaka, spotted green pufferfish, and Japanese pufferfish was prepared. To avoid the presence of annotated genes with different gene names or unannotated genes in the agastric genome data, the absence of the genes was confirmed using a BLAST search (TBLASTN)[90] of zebrafish, Japanese medaka, spotted green pufferfish, and Japanese pufferfish with Ensembl, and gene names with one or more orthologs were removed from the list. The presence of the orthologs of the listed genes for jawed vertebrate species listed Table 1 were analyzed by text search or TBLASTN analyses using Ensembl and NCBI. The synteny of each gene in the list was compared among the above species using Ensembl and NCBI.

Dot plot analysis

To analyze the pseudogenization or whole gene deletion of the eight genes *slc26a9*, *kcne2*, *vsig1*, *cldn18a*, *atp4a*, *atp4b*, *pga2*, and *pgc*, in the 11 agastric fish species, the coding region of each gene and its flanking regions of the gastric species, three-spined stickleback (*Gasterosteus aculeatus*), and channel catfish (*Ictalurus punctatus*) were compared with the corresponding genomic regions of the 11 agastric fish species listed in Fig. 1. Dot plot comparisons were performed using the EMBOSS dotmatcher program

Table 1 | Genome databases used for synteny analysis of gastric and agastric Actinopterygii species and the evolutionary analysis of *pga* in vertebrates

Category	Species	Genome database	Remarks
ray-finned fish	spotted gar (*Lepisosteus oculatus*)	GCF_000242695.1[100]	synteny analysis, evolution of *pga*
ray-finned fish	Asian arowana (*Scleropages formosus*)	GCF_900964775.1[101]	synteny analysis, evolution of *pga*
ray-finned fish	golden-line barbell (*Sinocyclocheilus grahami*)	GCF_001515645.1[102]	synteny analysis
ray-finned fish	zebrafish (*Danio rerio*)	GCF_000002035.6[17]	synteny analysis
ray-finned fish	fathead minnow (*Pimephales promelas*)	GCF_016745375.1[103]	synteny analysis
ray-finned fish	channel catfish (*Ictalurus punctatus*)	GCF_001660625.3[104]	synteny analysis, evolution of *pga*
ray-finned fish	Mexican tetra (*Astyanax mexicanus*)	GCF_023375975.1[105]	synteny analysis, evolution of *pga*
ray-finned fish	rainbow trout (*Oncorhynchus mykiss*)	GCF_013265735.2[106]	synteny analysis, evolution of *pga*
ray-finned fish	northern pike (*Esox lucius*)	GCF_011004845.1[107]	synteny analysis, evolution of *pga*
ray-finned fish	Atlantic cod (*Gadus morhua*)	GCF_902167405.1[23]	synteny analysis, evolution of *pga*
ray-finned fish	greater amberjack (*Seriola dumerili*)	GCF_002260705.1[108]	synteny analysis, evolution of *pga*
ray-finned fish	Japanese medaka (*Oryzias latipes*)	GCF_002234675.1[18]	synteny analysis
ray-finned fish	turquoise killifish (*Nothobranchius furzeri*)	GCF_027789165.1[109]	synteny analysis
ray-finned fish	platyfish (*Xiphophorus maculatus*)	GCF_002775205.1[110]	synteny analysis
ray-finned fish	Nile tilapia (*Oreochromis niloticus*)	GCF_001858045.2[24]	synteny analysis, evolution of *pga*
ray-finned fish	clown anemonefish (*Amphiprion ocellaris*)	GCF_022539595.1[111]	synteny analysis, evolution of *pga*
ray-finned fish	humphead wrasse (*Cheilinus undulatus*)	GCF_018320785.1[112]	synteny analysis
ray-finned fish	ballan wrasse (*Labrus bergylta*)	GCF_900080235.1[21]	synteny analysis
ray-finned fish	gilthead seabream (*Sparus aurata*)	GCF_900880675.1[113]	synteny analysis, evolution of *pga*
ray-finned fish	three-spined stickleback (*Gasterosteus aculeatus*)	GCF_016920845.1[22]	synteny analysis, evolution of *pga*
ray-finned fish	ocean sunfish (*Mola mola*)	GCA_001698575.1[114]	synteny analysis, evolution of *pga*
ray-finned fish	Japanese pufferfish (*Takifugu rubripes*)	GCF_901000725.2[19]	synteny analysis, evolution of *pga*
ray-finned fish	spotted green pufferfish (*Tetraodon nigroviridis*)	GCA_000180735.1[20]	synteny analysis, evolution of *pga*
lobe-finned fish	coelacanth (*Latimeria chalumnae*)	GCF_000225785.1[37]	synteny analysis, evolution of *pga*
tetrapod	tropical clawed frog (*Xenopus tropicalis*)	GCF_000004195.4[38]	synteny analysis, evolution of *pga*
tetrapod	anole lizard (*Anolis carolinensis*)	GCF_000090745.2[39]	synteny analysis, evolution of *pga*
tetrapod	platypus (*Ornithorhynchus anatinus*)	GCF_004115215.2[40]	synteny analysis
tetrapod	echidna (*Tachyglossus aculeatus*)	GCF_015852505.1[25]	synteny analysis
tetrapod	human (Homo sapiens)	GCF_000001405.40[41]	synteny analysis, evolution of *pga*
ray-finned fish	reedfish (*Erpetoichthys calabaricus*)	GCF_900747795.2	evolution of *pga*
ray-finned fish	gray bichir (*Polypterus senegalus*)	GCF_016835505.1[115]	evolution of *pga*
ray-finned fish	sterlet (*Acipenser ruthenus*)	GCF_902713425.1[116]	evolution of *pga*
ray-finned fish	Mississippi paddlefish (*Polyodon spathula*)	GCF_017654505.1[117]	evolution of *pga*
ray-finned fish	*Brienomyrus brachyistius*	GCF_023856365.1[118]	evolution of *pga*
ray-finned fish	*Paramormyrops kingsleyae*	GCF_002872115.1[119]	evolution of *pga*
ray-finned fish	Indo-pacific tarpon (*Megalops cyprinoides*)	GCF_013368585.1	evolution of *pga*
ray-finned fish	European eel (*Anguilla anguilla*)	GCF_013347855.1[120]	evolution of *pga*
ray-finned fish	Atlantic herring (*Clupea harengus*)	GCF_900700415.2[121]	evolution of *pga*
tetrapod	cow (*Bos taurus*)	GCF_002263795.3[122]	evolution of *pga*
tetrapod	chicken (*Gallus gallus*)	GCF_016699485.2[123]	evolution of *pga*
tetrapod	Japanese gecko (*Gekko japonicus*)	GCF_001447785.1[124]	evolution of *pga*
tetrapod	Reeves's turtle (*Mauremys reevesii*)	GCF_016161935.1[125]	evolution of *pga*
tetrapod	Congo dwarf clawed frog (*Hymenochirus boettgeri*)	GCA_019447015.1	evolution of *pga*
cartilaginous fish	great white shark (*Carcharodon carcharias*)	GCF_017639515.1[126]	evolution of *pga*
cartilaginous fish	whale shark (*Rhincodon typus*)	GCF_021869965.1[127]	evolution of *pga*
cartilaginous fish	zebra shark (*Stegostoma tigrinum*)	GCF_030684315.1[127]	evolution of *pga*
cartilaginous fish	whitespotted bambooshark (*Chiloscyllium plagiosum*)	GCF_004010195.1[128]	evolution of *pga*
cartilaginous fish	smaller spotted catshark (*Scyliorhinus canicula*)	GCF_902713615.1	evolution of *pga*
cartilaginous fish	thorny skate (*Amblyraja radiata*)	GCF_010909765.2	evolution of *pga*
cartilaginous fish	smalltooth sawfish (*Pristis pectinata*)	GCF_009764475.1	evolution of *pga*
cartilaginous fish	little skate (*Leucoraja erinacea*)	GCF_028641065.1[129]	evolution of *pga*

with a window size of 20 and threshold score of 70 (https://www.ebi.ac.uk/Tools/emboss/). To analyze the pseudogenization of platypus *vsig1*, a dot plot analysis was performed between echidna *vsig1* and its flanking regions and the corresponding genome regions of the platypus containing the *vsig1* pseudogene using the EMBOSS dotmatcher program with a window size of 20 and a threshold score of 70.

Phylogenetic and synteny analyses of *pga*

pga orthologs were identified in the genome data of ray-finned fish, lobe-finned fish, tetrapods, and cartilaginous fish, as listed in Table 1. The deduced amino acid sequences were aligned using ClustalW software, and a phylogenetic tree was constructed using MEGA11[91] using the maximum likelihood method. The synteny of *pga* was compared among the above species using the Ensembl and NCBI databases.

Semi-quantitative reverse transcription (RT)-PCR

Three-spined sticklebacks (*Gasterosteus aculeatus*) and humphead wrasses (*Cheilinus undulatus*) captured in Japan in 2012 and 2023, respectively, were obtained from local dealers. The animal protocols were in accordance with a manual approved by the Institutional Animal Experiment Committee of the Tokyo Institute of Technology. We have complied with all relevant ethical regulations for animal use. The fishes were anesthetized by immersion in 0.1% ethyl m-aminobenzoate methanesulfonate (MS222; Sigma, St. Louis, MO, USA), which was neutralized to pH 7.4 with sodium bicarbonate prior to use, and then decapitated. The tissues for RNA preparation were removed with ophthalmic scissors and frozen in liquid nitrogen. Tissues other than ovary and testis were once pooled without distinguishing between males and females. Ovary and testis were obtained from females and males, respectively, and pooled. Total RNA was isolated from the three-spined stickleback and humphead wrasse tissues by acid guanidinium thiocyanate-phenol-chloroform extraction using Isogen reagent (Nippon Gene, Tokyo, Japan) according to the manufacturer's manual. Owing to the small size of the three-spined sticklebacks, organs from three or more individuals were pooled for RNA extraction. Because only one 230-gram individual of humphead wrasse was available, RNA was extracted from organs derived from one individual. The RNA was dissolved in diethyl pyrocarbonate (DEPC)-treated water and its concentration was estimated by measuring the absorbance at 260 nm. mRNA preparations were reverse-transcribed into cDNA using the oligo(dT) primer and the SuperScript III First-Strand Synthesis System (Invitrogen). The cDNA (0.25 µL of the Super Script III reaction) was used as the template for PCRs, along with the specific primers shown in Supplementary Table 17. The PCR reactions were performed as follows[92]. Each reaction mixture (final volume, 12.5 µL) consisted of 0.25 µL cDNA (template), primers (individual final concentration, 0.25 µM), and 6.25 µL GoTaq Green Master Mix (2×; Promega, Madison, WI, USA). The PCR conditions were as follows: initial denaturation at 94 °C for 2 min, 28 or 33 cycles of 94 °C for 15 s (denaturation), 55 °C for 30 s (annealing), 72 °C for 1 min (extension), and a final extension at 72 °C for 7 min. PCR products from the three-spined sticklebacks were separated on agarose gels and visualized with ethidium bromide. The fluorescence images were analyzed with a Kodak Image Station 2000R system (Eastman Kodak, Rochester, NY, USA). The PCR products from the humphead wrasse were diluted and loaded onto a Microchip Electrophoresis system for DNA/RNA analysis (MCE-202 MultiNA; Shimadzu, Kyoto, Japan) using a DNA-12000 reagent kit (Shimadzu) following to the manufacturer's instructions. Electrophoresis results were analyzed using the MultiNA Viewer software (Shimadzu). Images of the gels are shown in Supplementary Fig. 11.

In situ hybridization histochemistry

In situ hybridization was performed as previously described in ref. 93. For tissue fixation, three-spined sticklebacks were anesthetized by immersion in 0.1% MS222, neutralized to pH 7.4, treated with sodium bicarbonate before use, and then decapitated. The stomach of three-spined sticklebacks was fixed in 4% paraformaldehyde in 0.1 M phosphate buffer at pH 7.4 for 1 d at 4 °C. Tissues were dehydrated, embedded in paraplast (Leica Microsystems, Wetzlar, Germany), and cut in 5 µm slices. For in situ hybridization, sections were deparaffinized in xylene, rehydrated by serial alcohol solutions, treated with proteinase K (5 µg/mL) for 10 min, and postfixed in 4% paraformaldehyde in 0.1 M phosphate buffer at pH 7.4. The sections were equilibrated in hybridization buffer (5× SSC and 50% formamide) at 58 °C for 2 h. A partial sequence of each target gene was cloned into the pGEM-T Easy vector (Promega) using the primers listed in Supplementary Table 17. Sense and antisense probes were prepared using a digoxigenin (DIG) RNA labeling kit (Roche Applied Science, Indianapolis, IN, USA), diluted in hybridization buffer containing calf thymus DNA (40 µg/mL), and denatured at 85 °C for 10 min. Denatured RNA probes were spread on the sections and incubated at 58 °C for >40 h depending on the expression level in a moist chamber saturated with hybridization buffer. Specific signals were developed using a DIG nucleic acid detection kit (Roche Applied Science), according to the manufacturer's protocol. Some sections were stained with hematoxylin and eosin (H&E) or periodic acid–Schiff to determine the basic structure of epithelial cells. Images were obtained using a TOCO automatic virtual slide system (Path Imaging, Tokyo, Japan) and a microscope equipped with a digital CCD camera (AxioCam HRc; Carl Zeiss, Oberkochen, Germany), and processed using AxioVision 4.1 software (Carl Zeiss).

Calculation of nucleotide substitution rates

Nucleotide sequences for claudin 18 were obtained from GenBank or Ensembl. We used three nucleotide sequences for tetrapod/coelacanth *cldn18* from human, tropical clawed frog, and coelacanth, three for each of gastric fish *cldn18a* and *cldn18b* from Atlantic cod, Nile tilapia, and three-spined stickleback, and two for agastric fish *cldn18b* from zebrafish and Japanese pufferfish. We used transcriptional sequences predicted from genome data when mRNA data was not available from the databases. The coding regions were aligned using ClustalW software[94] and sites containing gaps were deleted manually without shifting the reading frame (Supplementary Fig. 12). Distance values for the non-synonymous substitutions per site (dn) and synonymous substitutions per site (ds) were calculated based on the Nei-Gojobori (NG) method[95] using the alignment composed of 11 sequences and 522 positions and the MEGA6 software[96]. Standard errors were computed using the bootstrap method with 500 replicates. The number of non-synonymous differences (n), synonymous differences (s), non-synonymous sites (N), and synonymous sites (S) was calculated based on the Nei-Gojobori (NG) method using the MEGA6 software. Fisher's exact test was used for the statistical analyses[97].

Synteny analysis of monotremes and related species

The presence or absence of *atp4a*, *atp4b*, *pga*, *pgc*, and *vsig1* was confirmed by BLAST search (TBLASTN) and synteny analysis using the genome databases of coelacanth[37], *Xenopus*[38], anole lizard[39], platypus[40], echidna[25], and human[41]. Synteny analysis was performed manually using the Ensembl genome browser (https://www.ensembl.org)[98] or the NCBI genome data viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/)[99].

Statistics and reproducibility

All experiments using the three-spined stickleback and humphead wrasse were repeated at least twice, and reproducibility was confirmed using the same sample. For the statistical analyses of the of nucleotide substitution rates, we used three nucleotide sequences for tetrapod/coelacanth *cldn18*, three for each of gastric fish *cldn18a* and *cldn18b*, and two for agastric fish *cldn18b*. The numbers of sites for the statistical analyses are shown in Fig. 6d. Average numbers of non-synonymous differences (n) and unchanged non-synonymous sites (N-n) of gastric fish *cldn18b* were compared with those of agastric fishes, zebrafish, and Japanese pufferfish using by two-tailed Fisher's exact test using GraphPad Prism (GraphPad, San Diego, CA, USA) (https://www.graphpad.com/quickcalcs/contingency1/). Average numbers of synonymous differences (s) and unchanged synonymous sites (S-s) were also analyzed similarly by two-tailed Fisher's exact test.

Reporting summary

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

Data availability

All resources are available from the authors upon reasonable request.

Received: 3 April 2023; Accepted: 25 March 2024;
Published online: 03 April 2024

References

1. Near, T. J. et al. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. USA.* **109**, 13698–13703 (2012).
2. Nelson, J. S. *Fishes of the world*. 4th edn, (John Wiley and Sons, 2006).
3. Eschmeyer, W. N. & Fricke, R. *Catalog of Fishes*. Electronic Version edn, (California Academy of Sciences, 2012).
4. Castro, L. F. et al. Recurrent gene loss correlates with the evolution of stomach phenotypes in gnathostome history. *Proc. Biol. Sci.* **281**, 20132669 (2014).
5. Wilson, J. M. & Castro, L. F. Morphological diversity of the gastrointestinal tract in fishes. in *Fish Physiology* Vol. 30, 1–55 (Academic Press, 2010).
6. Protas, M., Conrad, M., Gross, J. B., Tabin, C. & Borowsky, R. Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. *Curr. Biol.* **17**, 452–454 (2007).
7. Romero, A. *Cave biology: life in darkness*. (Cambridge University Press, 2009).
8. Moran, D., Softley, R. & Warrant, E. J. Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism. *PLoS One* **9**, e107877 (2014).
9. Alexander, R. M. Buoyancy. in *The Physiology of Fishes* (ed D. H. Evans) 75–97 (CRC Press, 1993).
10. Hawkes, J. W. The structure of fish skin. I. General organization. *Cell Tissue Res.* **149**, 147–158 (1974).
11. Hickman, C. P. & Trump, B. F. The kidney. in *Fish Physiology. I*. (eds W. S. Hoar & D. J. Randall) 91–239 (Academic Press, 1969).
12. Cohn, M. J. & Tickle, C. Developmental basis of limblessness and axial patterning in snakes. *Nature* **399**, 474–479 (1999).
13. Skinner, A., Lee, M. S. & Hutchinson, M. N. Rapid and repeated limb loss in a clade of scincid lizards. *BMC Evol. Biol.* **8**, 310 (2008).
14. Thewissen, J. G. et al. Developmental basis for hind-limb loss in dolphins and origin of the cetacean bodyplan. *Proc. Natl. Acad. Sci. USA.* **103**, 8414–8418 (2006).
15. Bickford, D., Iskandar, D. & Barlian, A. A lungless frog discovered on Borneo. *Curr. Biol.* **18**, R374–375 (2008).
16. Ordonez, G. R. et al. Loss of genes implicated in gastric function during platypus evolution. *Genome Biol* **9**, R81 (2008).
17. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
18. Kasahara, M. et al. The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714–719 (2007).
19. Aparicio, S. et al. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
20. Jaillon, O. et al. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
21. Lie, K. K. et al. 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 (2018).
22. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
23. Star, B. et al. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207–210 (2011).
24. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
25. Zhou, Y. et al. Platypus and echidna genomes reveal mammalian biology and evolution. *Nature* **592**, 756–762 (2021).
26. Kato, A. et al. Identification of renal transporters involved in sulfate excretion in marine teleost fish. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **297**, R1647–1659 (2009).
27. Watanabe, T. & Takei, Y. Molecular physiology and functional morphology of SO_4^{2-} excretion by the kidney of seawater-adapted eels. *J. Exp. Biol.* **214**, 1783–1790 (2011).
28. Chang, M. H. et al. Slc26a9 - anion exchanger, channel and Na^+ transporter. *J. Membr. Biol.* **228**, 125–140 (2009).
29. Xu, J. et al. Deletion of the chloride transporter Slc26a9 causes loss of tubulovesicles in parietal cells and impairs acid secretion in the stomach. *Proc. Natl. Acad. Sci. USA.* **105**, 17955–17960 (2008).
30. Demitrack, E. S., Soleimani, M. & Montrose, M. H. Damage to the gastric epithelium activates cellular bicarbonate secretion via SLC26A9 Cl^-/HCO_3^-. *Am. J. Physiol. Gastrointest. Liver Physiol.* **299**, G255–264 (2010).
31. Anagnostopoulou, P. et al. SLC26A9-mediated chloride secretion prevents mucus obstruction in airway inflammation. *J. Clin. Invest.* **122**, 3629–3634 (2012).
32. Kinsella, R. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**, bar030 (2011).
33. Miura, Y., Suzuki-Matsubara, M., Kageyama, T. & Moriyama, A. Structure, molecular evolution, and hydrolytic specificities of largemouth bass pepsins. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* **192**, 49–59 (2016).
34. Nei, M. & Rooney, A. P. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* **39**, 121–152 (2005).
35. Meyer, A. & Van de Peer, Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**, 937–945 (2005).
36. Abe, K., Tani, K., Friedrich, T. & Fujiyoshi, Y. Cryo-EM structure of gastric H^+,K^+-ATPase with a single occupied cation-binding site. *Proc. Natl. Acad. Sci. USA.* **109**, 18401–18406 (2012).
37. Amemiya, C. T. et al. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311–316 (2013).
38. Hellsten, U. et al. The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633–636 (2010).
39. Alfoldi, J. et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011).
40. Pask, A. J. et al. Analysis of the platypus genome suggests a transposon origin for mammalian imprinting. *Genome Biol.* **10**, R1 (2009).
41. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
42. Albalat, R. & Canestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).
43. Blomme, T. et al. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43 (2006).
44. Brunet, F. G. et al. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol. Biol. Evol.* **23**, 1808–1816 (2006).
45. Canestro, C., Catchen, J. M., Rodriguez-Mari, A., Yokoi, H. & Postlethwait, J. H. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet.* **5**, e1000496 (2009).
46. Semon, M. & Wolfe, K. H. Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. *Trends Genet.* **23**, 108–112 (2007).
47. Wang, X., Grus, W. E. & Zhang, J. Gene losses during human origins. *PLoS Biol.* **4**, e52 (2006).

48. Hayashi, D. et al. Deficiency of claudin-18 causes paracellular H^+ leakage, up-regulation of interleukin-1beta, and atrophic gastritis in mice. *Gastroenterology* **142**, 292–304 (2012).
49. Oidovsambuu, O. et al. Adhesion protein VSIG1 is required for the proper differentiation of glandular gastric epithelia. *PLoS One* **6**, e25908 (2011).
50. Chanet, B. et al. Visceral anatomy of ocean sunfish (*Mola mola* (L., 1758), Molidae, Tetraodontiformes) and angler (*Lophius piscatorius* (L., 1758), Lophiidae, Lophiiformes) investigated by non-invasive imaging techniques. *C. R. Biol.* **335**, 744–752 (2012).
51. Fagundes, K. R., Rotundo, M. M. & Mari, R. B. Morphological and histochemical characterization of the digestive tract of the puffer fish *Sphoeroides testudineus* (Linnaeus 1758) (Tetraodontiformes: Tetraodontidae). *An. Acad. Bras. Cienc.* **88**, 1615–1624 (2016).
52. Kopic, S., Murek, M. & Geibel, J. P. Revisiting the parietal cell. *Am. J. Physiol. Cell Physiol.* **298**, C1–C10 (2010).
53. Schubert, M. L. Gastric secretion. *Curr. Opin. Gastroenterol.* **26**, 598–603 (2010).
54. Evans, D. H., Piermarini, P. M. & Choe, K. P. The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiol. Rev.* **85**, 97–177 (2005).
55. Hiroi, J. & McCormick, S. D. New insights into gill ionocyte and ion transporter function in euryhaline and diadromous fish. *Respir. Physiol. Neurobiol.* **184**, 257–268 (2012).
56. Wong, M. K., Pipil, S., Kato, A. & Takei, Y. Duplicated CFTR isoforms in eels diverged in regulatory structures and osmoregulatory functions. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **199**, 130–141 (2016).
57. Boettger, T. et al. Deafness and renal tubular acidosis in mice lacking the K-Cl co-transporter Kcc4. *Nature* **416**, 874–878 (2002).
58. Cowley, E. A. & Linsdell, P. Characterization of basolateral K^+ channels underlying anion secretion in the human airway cell line Calu-3. *J. Physiol.* **538**, 747–757 (2002).
59. Soler Artigas, M. et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* **43**, 1082–1090 (2011).
60. Bertrand, C. A., Zhang, R., Pilewski, J. M. & Frizzell, R. A. SLC26A9 is a constitutively active, CFTR-regulated anion conductance in human bronchial epithelia. *J. Gen. Physiol.* **133**, 421–438 (2009).
61. Chang, M. H. et al. Slc26a9 is inhibited by the R-region of the cystic fibrosis transmembrane conductance regulator via the STAS domain. *J. Biol. Chem.* **284**, 28306–28318 (2009).
62. LaFemina, M. J. et al. Claudin-18 deficiency results in alveolar barrier dysfunction and impaired alveologenesis in mice. *Am. J. Respir. Cell Mol. Biol.* **51**, 550–558 (2014).
63. Niimi, T. et al. claudin-18, a novel downstream target gene for the T/EBP/NKX2.1 homeodomain transcription factor, encodes lung- and stomach-specific isoforms through alternative splicing. *Mol. Cell. Biol.* **21**, 7380–7390 (2001).
64. Strug, L. J. et al. Cystic fibrosis gene modifier SLC26A9 modulates airway response to CFTR-directed therapeutics. *Hum. Mol. Genet.* **25**, 4590–4600 (2016).
65. Gong, J. et al. Genetic evidence supports the development of SLC26A9 targeting therapies for the treatment of lung disease. *NPJ Genom. Med.* **7**, 28 (2022).
66. Sun, L. et al. Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. *Nat. Genet.* **44**, 562–569 (2012).
67. Scanlan, M. J. et al. Glycoprotein A34, a novel target for antibody-based cancer immunotherapy. *Cancer Immun.* **6**, 2 (2006).
68. Bernal, C. et al. Functional Pro-metastatic Heterogeneity Revealed by Spiked-scRNAseq Is Shaped by Cancer Cell Interactions and Restricted by VSIG1. *Cell Rep.* **33**, 108372 (2020).
69. Zanconato, F., Cordenonsi, M. & Piccolo, S. YAP/TAZ at the Roots of Cancer. *Cancer Cell* **29**, 783–803 (2016).
70. Zhao, B., Li, L., Lei, Q. & Guan, K. L. The Hippo-YAP pathway in organ size control and tumorigenesis: an updated version. *Genes Dev.* **24**, 862–874 (2010).
71. Loh, Y. H., Christoffels, A., Brenner, S., Hunziker, W. & Venkatesh, B. Extensive expansion of the claudin gene family in the teleost fish, Fugu rubripes. *Genome Res.* **14**, 1248–1257 (2004).
72. Baltzegar, D. A., Reading, B. J., Brune, E. S. & Borski, R. J. Phylogenetic revision of the claudin gene family. *Mar. Genomics* **11**, 17–26 (2013).
73. Hou, J. et al. Claudin-16 and claudin-19 interaction is required for their assembly into tight junctions and for renal reabsorption of magnesium. *Proc. Natl. Acad. Sci. USA.* **106**, 15350–15355 (2009).
74. Meoli, L. & Gunzel, D. The role of claudins in homeostasis. *Nat. Rev. Nephrol.* **19**, 587–603 (2023).
75. Yu, A. S. Claudins and the kidney. *J. Am. Soc. Nephrol.* **26**, 11–19 (2015).
76. Kurokawa, T., Uji, S. & Suzuki, T. Identification of pepsinogen gene in the genome of stomachless fish, *Takifugu rubripes*. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* **140**, 133–140 (2005).
77. Erawijantari, P. P. et al. Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. *Gut* **69**, 1404–1415 (2020).
78. Cahill, M. M. Bacterial flora of fishes: A review. *Microb. Ecol.* **19**, 21–41 (1990).
79. Tarnecki, A. M., Burgos, F. A., Ray, C. L. & Arias, C. R. Fish intestinal microbiome: diversity and symbiosis unravelled by metagenomics. *J. Appl. Microbiol.* **123**, 2–17 (2017).
80. Clements, K. D., Angert, E. R., Montgomery, W. L. & Choat, J. H. Intestinal microbiota in fishes: what's known and what's not. *Mol. Ecol.* **23**, 1891–1898 (2014).
81. Egerton, S., Culloty, S., Whooley, J., Stanton, C. & Ross, R. P. The Gut Microbiota of Marine Fish. *Front. Microbiol.* **9**, 873 (2018).
82. German, D. P. Inside the guts of wood-eating catfishes: can they digest wood? *J. Comp. Physiol. B* **179**, 1011–1023 (2009).
83. Roeselers, G. et al. Evidence for a core gut microbiota in the zebrafish. *ISME J* **5**, 1595–1608 (2011).
84. Sullam, K. E. et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol. Ecol.* **21**, 3363–3378 (2012).
85. Leigh, S. C., Catabay, C. & German, D. P. Sustained changes in digestive physiology and microbiome across sequential generations of zebrafish fed different diets. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **273**, 111285 (2022).
86. Sakamoto, N. et al. Role for cGATA-5 in transcriptional regulation of the embryonic chicken pepsinogen gene by epithelial-mesenchymal interactions in the developing chicken stomach. *Dev. Biol.* **223**, 103–113 (2000).
87. Shin, M., Noji, S., Neubuser, A. & Yasugi, S. FGF10 is required for cell proliferation and gland formation in the stomach epithelium of the chicken embryo. *Dev. Biol.* **294**, 11–23 (2006).
88. Spencer-Dene, B. et al. Stomach development is dependent on fibroblast growth factor 10/fibroblast growth factor receptor 2b-mediated signaling. *Gastroenterology* **130**, 1233–1244 (2006).
89. Cunningham, F. et al. Ensembl 2015. *Nucleic Acids Res.* **43**, D662–669 (2015).
90. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
91. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
92. Tran, Y. H. et al. Spliced isoforms of LIM-domain-binding protein (CLIM/NLI/Ldb) lacking the LIM-interaction domain. *J. Biochem.* **140**, 105–119 (2006).

93. Takei, Y. et al. Molecular mechanisms underlying active desalination and low water permeability in the esophagus of eels acclimated to seawater. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **312**, R231–R244 (2017).
94. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
95. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
96. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
97. Zhang, J., Rosenberg, H. F. & Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA.* **95**, 3708–3713 (1998).
98. Martin, F. J. et al. Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).
99. Rangwala, S. H. et al. Accessing NCBI data using the NCBI Sequence Viewer and Genome Data Viewer (GDV). *Genome Res.* **31**, 159–169 (2021).
100. Braasch, I. et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **48**, 427–437 (2016).
101. Bian, C. et al. The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Sci. Rep.* **6**, 24501 (2016).
102. Yang, J. et al. The Sinocyclocheilus cavefish genome provides insights into cave adaptation. *BMC Biol.* **14**, 1 (2016).
103. Burns, F. R. et al. Sequencing and de novo draft assemblies of a fathead minnow (*Pimephales promelas*) reference genome. *Environ. Toxicol. Chem.* **35**, 212–217 (2016).
104. Chen, X. et al. High-quality genome assembly of channel catfish, *Ictalurus punctatus*. *Gigascience* **5**, 39 (2016).
105. Warren, W. C. et al. A chromosome-level genome of *Astyanax mexicanus* surface fish for comparing population-specific genetic differences contributing to trait evolution. *Nat. Commun.* **12**, 1447 (2021).
106. Palti, Y. et al. A second generation integrated map of the rainbow trout (*Oncorhynchus mykiss*) genome: analysis of conserved synteny with model fish genomes. *Mar Biotechnol (NY)* **14**, 343–357 (2012).
107. Rondeau, E. B. et al. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. *PLoS One* **9**, e102089 (2014).
108. Araki, K. et al. Whole Genome Sequencing of Greater Amberjack (*Seriola dumerili*) for SNP Identification on Aligned Scaffolds and Genome Structural Variation Analysis Using Parallel Resequencing. *Int. J. Genomics* **2018**, 7984292 (2018).
109. Reichwald, K. et al. High tandem repeat content in the genome of the short-lived annual fish *Nothobranchius furzeri*: a new vertebrate model for aging research. *Genome Biol* **10**, R16 (2009).
110. Schartl, M. et al. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat. Genet.* **45**, 567–572 (2013).
111. Tan, M. H. et al. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *Gigascience* **7**, 1–6 (2018).
112. Liu, D. et al. Chromosome-level genome assembly of the endangered humphead wrasse *Cheilinus undulatus*: Insight into the expansion of opsin genes in fishes. *Mol. Ecol. Resour.* **21**, 2388–2406 (2021).
113. Pauletto, M. et al. Genomic analysis of *Sparus aurata* reveals the evolutionary dynamics of sex-biased genes in a sequential hermaphrodite fish. *Commun. Biol.* **1**, 119 (2018).
114. Pan, H. et al. The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate. *Gigascience* **5**, s13742-13016-10144-13743, https://doi.org/10.1186/s13742-016-0144-3 (2016).
115. Bi, X. et al. Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes. *Cell* **184**, 1377–1391e1314 (2021).
116. Cheng, P. et al. Draft Genome and Complete Hox-Cluster Characterization of the Sterlet (*Acipenser ruthenus*). *Front. Genet.* **10**, 776 (2019).
117. Cheng, P. et al. The American Paddlefish Genome Provides Novel Insights into Chromosomal Evolution and Bone Mineralization in Early Vertebrates. *Mol. Biol. Evol.* **38**, 1595–1607 (2021).
118. Peterson, R., Sullivan, J. & Pirro, S. The Complete Genome Sequences of 38 Species of Elephantfishes (Mormyridae, Osteoglossiformes). *Biodivers. Genomes* **2022**, https://doi.org/10.56179/001c.56077 (2022).
119. Gallant, J. R., Losilla, M., Tomlinson, C. & Warren, W. C. The Genome and Adult Somatic Transcriptome of the Mormyrid Electric Fish Paramormyrops kingsleyae. *Genome Biol. Evol.* **9**, 3525–3530 (2017).
120. Coppe, A. et al. Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel. *BMC Genomics* **11**, 635 (2010).
121. Pettersson, M. E. et al. A Long-Standing Hybrid Population Between Pacific and Atlantic Herring in a Subarctic Fjord of Norway. *Genome Biol. Evol*. **15**, https://doi.org/10.1093/gbe/evad069 (2023).
122. Zimin, A. V. et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **10**, R42 (2009).
123. International Chicken Genome Sequencing, C. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
124. Liu, Y. et al. *Gekko japonicus* genome reveals evolution of adhesive toe pads and tail regeneration. *Nat. Commun.* **6**, 10033 (2015).
125. Liu, J. et al. Chromosome-level genome assembly of the Chinese three-keeled pond turtle (*Mauremys reevesii*) provides insights into freshwater adaptation. *Mol. Ecol. Resour.* **22**, 1596–1605 (2022).
126. Marra, N. J. et al. White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proc. Natl. Acad. Sci. USA.* **116**, 4446–4455 (2019).
127. Yamaguchi, K. et al. Elasmobranch genome sequencing reveals evolutionary trends of vertebrate karyotype organization. *Genome Res.* **33**, 1527–1540 (2023).
128. Zhao, R. et al. Genomic Comparison and Genetic Marker Identification of the White-Spotted Bamboo Shark *Chiloscyllium plagiosum*. *Front. Mar. Sci.* **9**, 936681 (2022).
129. Marletaz, F. et al. The little skate genome and the evolutionary emergence of wing-like fins. *Nature* **616**, 495–503 (2023).
130. Near, T. J. et al. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl. Acad. Sci. USA.* **110**, 12738–12743 (2013).
131. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
132. Grahammer, F. et al. The cardiac K^+ channel KCNQ1 is essential for gastric acid secretion. *Gastroenterology* **120**, 1363–1371 (2001).
133. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).

Acknowledgements
We thank Yoko Yamamoto, Nana Shinohara, and the Biomaterials Analysis Division at the Tokyo Institute of Technology for technical assistance; Megumi Ohmaki and Yuko Akiyoshi for their secretarial assistance; and the anonymous reviewers for their useful comments. This work was supported by Japan Society for the Promotion of Science KAKENHI (Grants 24651211, 26292113, and 21H02281). The Romero laboratory work was supported by NIH R01-EY017732, R21-DK129897, and R01-DK128844.

Author contributions
A.K., T.W., M.F.R. and Y.T. conceived the study; A.K., S.P., M.Ku, T.W., A.P.C., Z.I., N.H., M.W., M.Ko, M.F.R. and Y.T. designed and conducted the expression analyses; A.K., C.O., T.W. and A.N. designed and conducted the bioinformatics analyses; A.K. and Y.T. wrote the paper; and A.K., C.O. and A.N. prepared the revised manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06103-x.

Correspondence and requests for materials should be addressed to Akira Kato.

Peer review information *Communications Biology* thanks Donovan German and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Luke R. Grinham.