Evolution of the Jawed Vertebrate (Gnathostomata) Stomach through Gene Repertoire Loss: Findings from Agastric Species

Jackson Dann[a], Frank Grützner[ab*]

[a]School of Biological Sciences, The University of Adelaide, SA, 5005, Australia

[b] Robinson Research Institute, The University of Adelaide, SA, 5005, Australia

*Correspondence to frank.grutzner@adelaide.edu.au

Abstract

Independent and convergent gene loss, such as the loss of gastric enzymes (*Pga*, *Pgc*) and proton pump subunits (*Atp4a*, *Atp4b*) in monotremes and polyphyletic teleost species have been associated with the agastric phenotype – stomachs lacking gross morphology, gastric glands and acid secretion. The loss of the key gastric hormone gastrin (*Gast*) has been observed in monotremes, but not corroborated in agastric teleosts. Furthermore, it is unclear how this loss affects the evolution and selection of the native receptor (*Cckbr*), gastrin releasing peptide (*Grp*) and receptor (*Grpr*), or if there are further gene losses associated with the agastric phenotype. We investigated the evolution, selection and conservation of the gastrin physiological pathway and novel gastric genes in a broad range of gastric and agastric vertebrates. Our analyses showed no correlation between the losses of *Gast* and *Grpr* with the agastric phenotype in vertebrate lineages. We identified genes implicated in gastric mucosal maintenance and development (*Gkn1*, *Gkn2*, *Tff1*, *Tff2*, *Vsig1*, *Anxa10*) that are lost in agastric species except the echidna, which retained some of them (*Gkn1*, *Tff2*, *Vsig1*). Our findings demonstrate a convergent repertoire of gastric genes lost in agastric species, partial retention of gastric functionality in the echidna, and the diversified roles of gastrin pathway constituents in vertebrate lineages.

Introduction

The stomach is a highly conserved and integral organ throughout jawed vertebrate (Gnathostomata) evolution as evidenced by conserved developmental pathways dating back to protochordate species (Nakazawa et al., 2013; Nakayama et al., 2017). Though morphological features can vary widely between taxa, comparable components include storage compartments (enlargement of the lumen) and acid-secreting glandular epithelium (Koelz, 1992; Karasov, Douglas & Ronald, 2013). Despite gastric features being nearly ubiquitous, a unique gastric phenotype has evolved over 15 times independently in ray-finned fish (Actinopterygii), and at least twice in lobe-finned fish (Sarcopterygii): once in lungfish and once in monotremes. This gastric phenotype – hereby known as agastric – is typified by an absence of acidic luminal contents, loss of glandular epithelium and a significant reduction in overall size and gross morphology (Wilson & Castro, 2010; Castro et al., 2014).

Not only is the agastric phenotype convergent morphologically between species, but agastric species share the concomitant loss of genes for gastric enzymes (*Pga*, *Pgc*) and proton pump subunits (*Atp4a*, *Atp4b*). In addition to this, monotremes have had an inactivation of the NK3 homeobox 2 (*Nkx3.2*) and have lost gastrin (*Gast*): a key gastric hormone regulated by the upstream neuropeptide gastrin-releasing peptide (*Grp)* and its native receptor (*Grpr*), which when released and binds the native cholecystokinin B receptor (*Cckbr*) potentiates gastric acid secretion (Ordoñez et al., 2008; Schubert & Peura, 2008; Castro et al., 2014; Zhou et al., 2021; Dann et al., 2024). Given the heterogeneity of the gastric epithelium in cell type (e.g. parietal cell, mucous cells, stem cells) and function (e.g. acidic digestion, immunity, hormone release), and various physiological and intracellular signalling pathways regulating gastric development and function, many more genes are expected to be affected by this drastic shift in gastric anatomy and physiology (Kim & Shivdasani, 2016; Agace & McCoy, 2017).

Recent discoveries in vertebrate developmental genetics and genomics have expanded the repertoire of genes implicated in gastric development and function. These have included the constituents of the gastrokine and trefoil factor families (*Gkn1*, *Gkn2*, *Gkn3*, *Tff1*, *Tff2*); secreted products of gastric mucosal epithelia with both independent and synergistic anti-inflammatory, tumour-suppressing and anti-apoptotic roles (Menheniott, Kurklu & Giraud, 2013; Hoffman, 2020). As well, *Vsig1* – a junction adhesion molecule expressed highly in gastric epithelia – and annexin A10 (*Anxa10*) – a component of the phospholipid-binding protein annexin superfamily – have displayed restricted gastric epithelial expression and been implicated in squamous versus glandular epithelial differentiation (Lu et al., 2011; Oidovsambuu et al., 2011). Despite some extensive work on expression patterns (particularly in gastrokine and trefoil factor families), and a recent work by Kato et al., (2024) – which noted the absence of *Vsig1* in agastric teleost lineages and the platypus – it is still currently unclear what emergent properties (i.e. gene ontologies) arise from the list of lost genes, and whether these additional genes are conserved in agastric vertebrate lineages (Jiang, Lossie & Applegate, 2011; Geahlen et al., 2013).

To better understand the evolution of the agastric phenotype in the Actinopterygii and Sarcopterygii clades, we explored the evolution, selection and loss of the gastrin physiological pathway along with novel gastric gene sets in the genomes of gastric and agastric taxa.

Materials and Methods

Sequence retrieval and synteny analysis

Sequences were obtained from evolutionary representative taxa of tetrapods (placental mammals, marsupials, monotremes, reptiles), Sarcopterygii (lobe-finned fish) and Actinopterygii (ray-finned fish). Annotated protein sequences were downloaded from NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and, where no annotations were present, searched for through the BLAST tool using homologous genes from closely-related species (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Synteny analysis was performed in the NCBI genome data viewer tool (https://www.ncbi.nlm.nih.gov/genome/gdv/). See Supplementary file 1 for complete sequence and genome information.

Sequence alignments and phylogenies

Multiple sequence alignments were performed using the Clustal Omega plugin in Geneious 11.0.14+1 except for deleted / pseudogenised monotreme genes (Geneious, 2022; Sievers et al., 2011). Protein maximum likelihood phylogenies were then constructed using the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) tree inference tool with standard settings and 1000 bootstrap replicates (Trifinopoulos et al., 2016). Substitution models derived from Bayesian information criterion and alignment outgroups can be found in Table S1.

Selection analysis (Ka/Ks ratios)

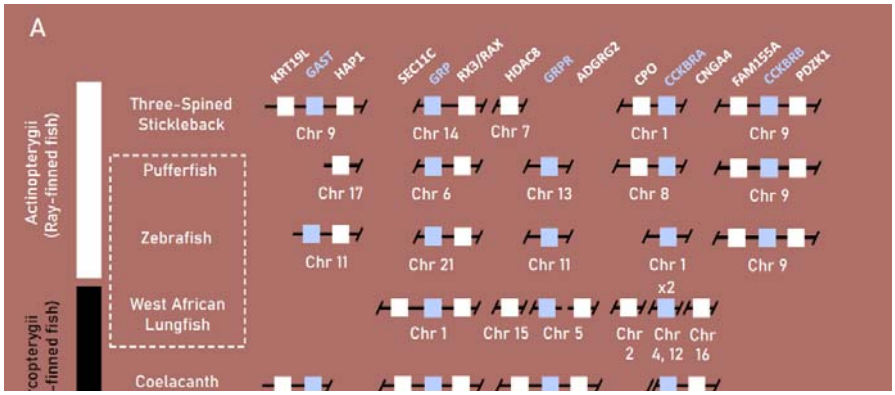
To detect signatures of selection, amino acid alignments and corresponding open reading frame DNA alignments were converted into PAML format codon alignments using PAL2NAL v 14 (Mikita, Torrents & Bork, 2006). Codon alignments along with their corresponding phylogenies in Newick format were then input into EasyCodeML v1.41 using the preset branch model with multiples models of foreground branch selection for each

lineage and likelihood ratio tests (LRTs) to assess model significance (Gao et al., 2019). EasyCodeML outputs can be found in Table S2.

Results

Conservation and loss of the gastrin release pathway and novel gastric gene repertoire in gastric and agastric jawed vertebrates (Gnathostomata)

Synteny analysis of gastrin (*Gast*), the native gastrin receptor (*Cckbr*), the gastrin-releasing peptide (*Grp*) and its receptor (*Grpr*) shows conserved syntenic blocks in ray- and lobe-finned fish species with recurring differences of flanking loci between lineages (Figure 1A). Gastrin was lost in two of the three agastric species (pufferfish and west African lungfish), but conserved in the agastric zebrafish. Additionally, the gastrin-releasing peptide receptor (*Grpr*) was absent from conserved genomic neighbourhoods in the gastric three-spined stickleback (*Gasterosteus aculeatus*). The highly-conserved syntenic blocks of tetrapod species corroborated the previously-noted gene losses of gastrin in both monotreme species (Ordoñez et al., 2016), with our analyses outlining an unexpected absence of the *Grpr* locus in the platypus (Figure 1B). Interestingly, these findings suggest that the loss of gastrin does not correlate with the presence of the agastric phenotype, and, that the loss of gastrin does not correlate with the loss of *Grpr*.



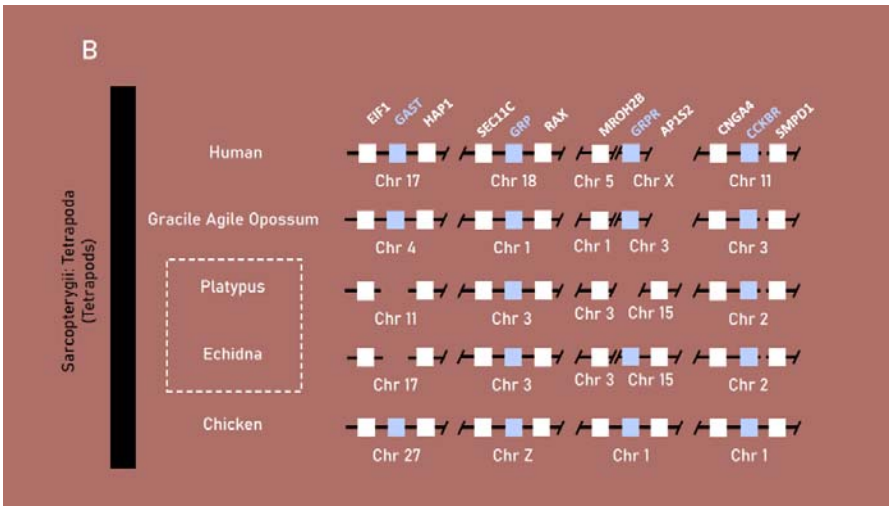


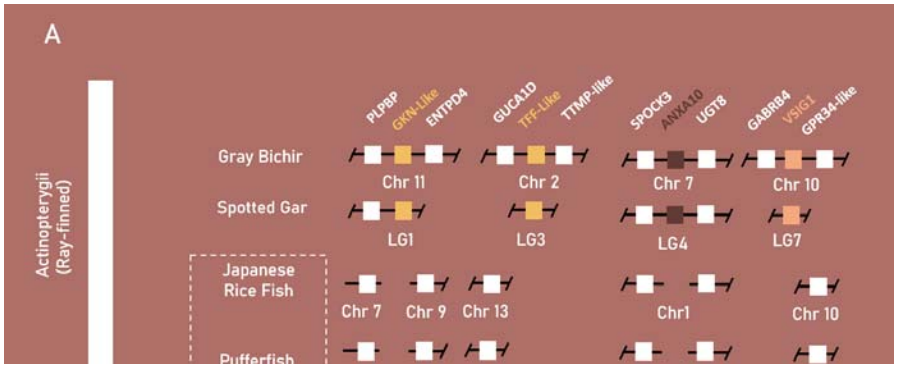
Figure 1: Synteny plots of gastrin (*Gast*), gastrin-releasing peptide (*Grp*), gastrin-releasing peptide receptor (*Grpr*) and cholecystokinin receptor B (*Cckbr*) in select vertebrate species. **A**, synteny plot of gastric genes lost or pseudogenised in the agastric species of the zebrafish (*D. rerio*) and West-African Lungfish (*P. annectens*) compared to the gastric three-spined stickleback (*G. aculeatus*) and coelacanth (*L. chalumnae*). **B,** Synteny plot of the short-beaked echidna (*T. aculeatus*) and platypus (*O. anatinus*) when compared to humans (*H. sapiens* / placental mammals), gracile agile opossums (*G. agilis* / marsupials) and chickens (*G. gallus* / reptiles). Class-level taxonomy (Sarcopterygii vs Actinopterygii) is displayed on the left, boxes with dotted lines indicates agastric species, coloured boxes indicate genes of interest, black boxes indicate pseudogene presence, numbers within a box indicates presence of only that family member and breaks or missing boxes in the syntenic block indicate gene absence.

We then explored the conservation and loss of genes implicated in gastric epithelial maintenance and development in jawed vertebrates. BLAST searching and synteny analysis of the sequences bearing gastrokine family homology (*Gkn1*, *Gkn2*, *Gkn3*), trefoil factor

homology (*Tff1*, *Tff2*), as well as *Anxa10* and *Vsig1* indicated they were lost in agastric Actinopterygii taxa but conserved in the gastric grey bichir and spotted gar (Figure 2A).

In Sarcopterygii, gastrokine family paralogs (*Gkn1*, *Gkn2*, *Gkn3*) were mainly conserved together within a syntenic block flanked by *Antxr1* and *Bmp10*. Gastrokine paralogs 2 and 3 were lost in both monotreme species, but *Gkn1* was retained in the echidna. The three Gastrokine paralogs found in the West-African lungfish did not place within the conserved genomic region, and, further phylogenetic analysis of these sequences found that basal Sarcopterygii sequences (coelacanth, west-African lungfish) placed in a separate clade to all three vertebrate paralogs. Furthermore, the echidna *Gkn1* sequence displayed higher similarity to *Gkn3* sequences, and was placed within the vertebrate *Gkn3* clade with 100 percent bootstrap support (Figure S1).

Trefoil factor paralogs 1 and 2 (*Tff1*, *Tff2*) were absent in the coelacanth and West-African lungfish and found only in tetrapod species in a highly conserved syntenic block between *Tff3* and *Spock3* except in the platypus – missing both – and the echidna, which was missing *Tff1*. Genomic sequences of *Anxa10* and *Vsig1 w*ere found in all Sarcopterygii taxa but not in the platypus and west-African lungfish, confirming the association with the agastric phenotype found by Kato et al., (2024), (Figure 2B). These findings indicate that these genes evolved prior to the divergence of Actinopterygii and Sactoperygii, and were subsequently lost in agastric species.



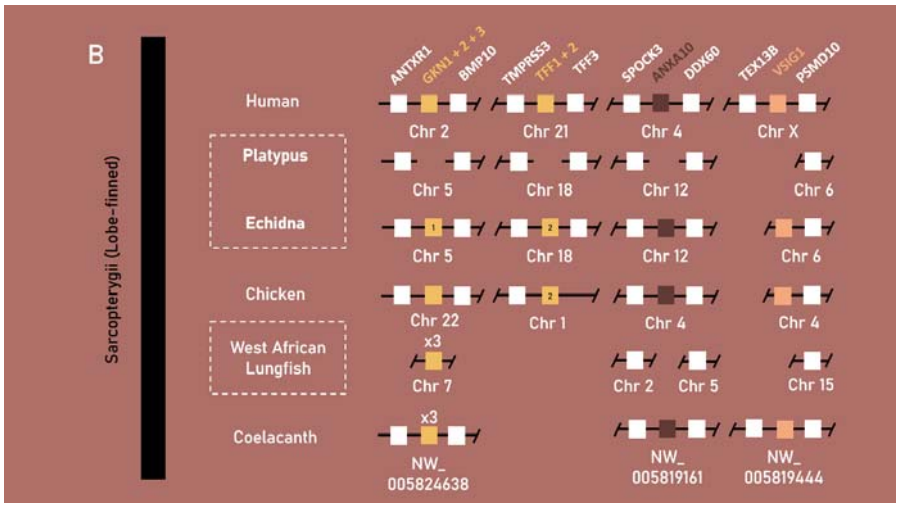


Figure 2: Synteny plots of genes implicated in gastric mucosal maintenance (*Gkn1*, *Gkn2*), gastric development (*Vsig1*) and unclear roles in gastric mucosal maintenance, growth and differentiation (*Anxa10*) in select vertebrate species. **A**, Synteny plot of Actinopterygii taxa: the agastric teleost species of the Japanese rice fish (*O. latipes*), pufferfish (*T. rubripes*) and zebrafish (*D. rerio*) compared to the gastric grey bichir (*P. senegalus*) and spotted gar (*L. oculatus*). **B** synteny plot of Sarcopterygii taxa: the agastric short-beaked echidna (*T. aculeatus*), platypus (*O. anatinus*) and West-African lungfish (*P. annectens*) compared to the gastric human (*H. sapien*), chicken (*G. gallus*) and coelacanth (*L chalumnae*). Class-level taxonomy (Sarcopterygii vs Actinopterygii) is displayed on the left, boxes with dotted lines indicates agastric species, coloured boxes indicate genes of interest, black boxes indicate pseudogene presence, numbers within a box indicates presence of only that family member and breaks or missing boxes in the syntenic block indicate gene absence.

Sequence evolution and selection pressures in the gastrin release pathway between gastric and agastric jawed vertebrates (Gnathostomata)

We then investigated sequence evolution and Ka/Ks ratios in the gastrin release pathway in gastric and agastric vertebrates to see whether gastrin loss affected the evolution or selection of other pathway components.

The Actinopterygii clade of gastrin sequences displayed minimal difference in substitution per site between the gastric three-spined stickleback (0.48) and agastric Japanese rice fish (0.34),(Figure 3A). Branches leading to monotreme and therian mammal CCKBR clades displayed identical lengths (0.17 substitutions per site), but differed more post-divergence into the mouse and wombat (0.41 and 0.25 substitutions per site) when compared to the agastric platypus and echidna (0.07 and $2x10^{-6}$ substitutions per site). Actinopterygii CCKBRA sequences had similar substitutions per site within the clade of the pufferfish (0.15), the Japanese rice fish (0.15) and the three-spined stickleback (0.12). The pufferfish CCKBRB sequence contained a larger number of substitutions per site (0.2) when compared to three-spined stickleback (0.07) and Japanese rice fish (0.13), (Figure 3B).

In the Sarcopterygii clade, *Grp* sequences from the coelacanth and west-African lungfish displayed large variations in substitutions per site (0.2 vs 0.6). Though echidna and platypus sequences displayed identical estimated substitutions per site (0.06), there was substantial difference in branch length to the root of mammalian sequences between monotremes (0.16), marsupials (0.44) and placental mammals (0.7). In the Actinopterygii clade, branch lengths differed to the root of the clade with the Japanese rice fish (0.76), pufferfish (0.41) and three-spined stickleback (0.55), (Figure 3C). Sarcopterygii and Actinopterygii *Grpr* sequences little difference in substitutions per site between gastric and agastric species, as shown by comparisons between the salmon (0.15) and the pufferfish (0.18), and comparisons of the coelacanth (0.08) and west-African lungfish (0.07), (Figure 3D). These findings suggest acceleration of sequence evolution from gastrin release pathway constituents independent of gastric phenotypes (gastric versus agastric).

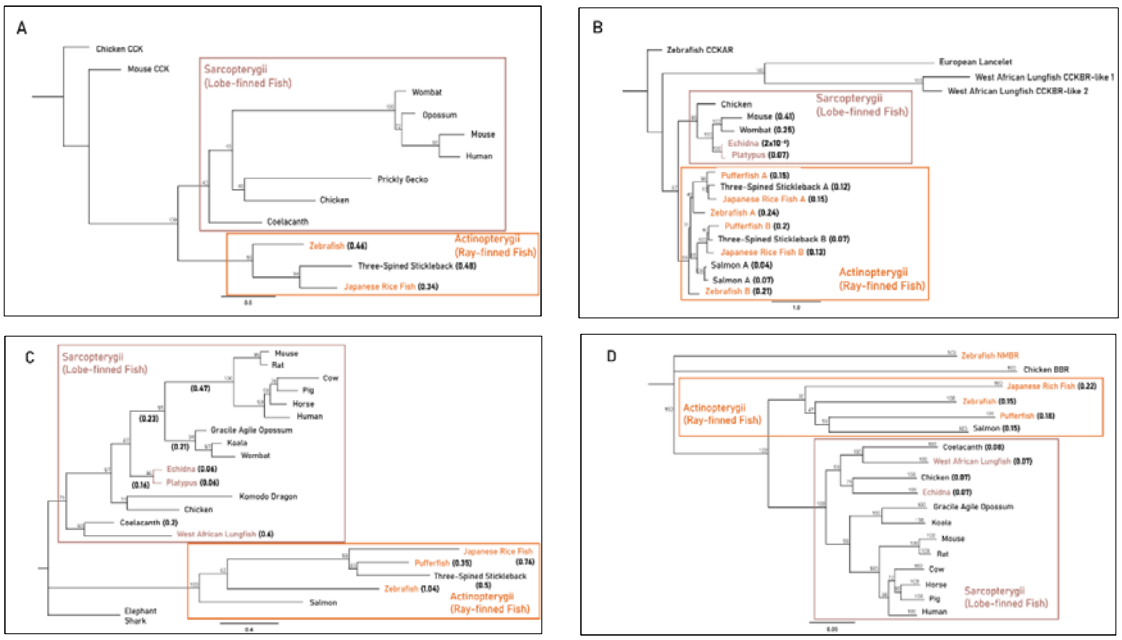


Figure 3: Maximum-likelihood amino acid phylogenies of the gastrin release pathway constituents: **A,** gastrin (GAST), **B,** cholecystokinin B receptor (CCKBR), **C,** gastrin-releasing peptide (GRP) and **D,** gastrin-releasing peptide receptor (GRPR) for select gastric and agastric vertebrate species constructed by the IQ-TREE web server with 1000 bootstrap replicates. Taxa, substitution models as chosen by Bayesian information criterion (BIC), outgroups and sequence references are in supplementary file 1. Paralogues of the CCKBR receptor for teleost lineages are identified with the letters A and B. Branch labels denote bootstrap percentage (%), coloured tip labels represent agastric lineages, coloured boxes outline class-level taxonomic groupings and branch length, scale bar and bracketed numbers measure substitutions per site to the nearest bifurcation.

Finally, we calculated the ratio of synonymous to non-synonymous substitutions (Ka/Ks) using maximum-likelihood branch model selection analyses. We found overall that the *Gast*, *Cckbr*, *Grp* and *Grpr* genes were under purifying selection (Table S1). Foreground scenarios with significant likelihood-ratio tests (LRTs) in the *Cckbr* phylogeny included branches leading to the clade with Japanese rice fish, pufferfish and three-spined stickleback *Cckbrb* (*p*

= 0.04, ω1 = 0.19, ω0 = 0.08), the three-spined stickleback *Cckbra* (p = 0.01, ω1 = $3x10^{-3,}$ ω0 = 0.07) and *Cckbrb* sequences (p = $1.12x10^{-5}$, ω1 = 0.33, ω0 = 0.08). In the *Grp* phylogeny significant LRTs included branches leading to the Tetrapod clade (p = $2.17x10^{-5}$, ω1 = $8.9x10^{-4}$, ω0 = 0.28), the mammalian clade (p = 0.03, ω1 = 0.02, ω0 = 0.24) and the mouse

Gene	Foreground branch	LRT *p* value	Foreground Ka/Ks (ω1)	Background Ka/Ks (ω0)

sequence (p = 0.04, ω1 = 0.02, , ω0 = 0.24). The *Grpr* phylogeny included two significant LRTs; the branch leading to the pufferfish sequence (p = 0.02, ω1 = 0.11, ω0 = 0.04) and to the Sarcopterygii clade (p = $8.93x10^{-7}$, ω1 = $1.94x10^{-3}$, ω0 = 0.05), (Table 1). These findings indicate that independent bursts of sequence evolution in gastrin release pathway constituents consisted largely of conservative mutations with no direct correlation to gastric phenotype.

Table 1: Summary of significant findings from CodeML branch model selection analysis of *Cckbr*, *Grp* and *Grpr* sequences with columns for foreground branch (taxonomic grouping or species), the model-fitting metric (likelihood-ratio test p) and Ka/Ks ratio under differing foreground scenarios (foreground = ω1, background = ω0).

Cckbr	Japanese Rice Fish B, Pufferfish B and Three-Spined Stickleback B	0.04	0.19	0.08
Cckbr	Three-Spined Stickleback B	$1.12x10^{-5}$	0.33	0.07
Cckbr	Three-Spined Stickleback A	0.01	$3x10^{-3}$	0.08
Grp	Tetrapoda (Mammals and Reptiles)	$2.17x10^{-5}$	$8.9x10^{-4}$	0.28
Grp	Mammalia	0.03	0.02	0.24
Grp	Mouse	0.04	0.06	0.24
Grpr	Pufferfish	0.02	0.11	0.04
Grpr	Sarcopterygii (Mammals, Reptiles and Lobe-Finned Fish)	$8.93x10^{-7}$	$1.94x10^{-3}$	0.05

Discussion

Stomach anatomy in vertebrates can been broadly described as gastric or agastric. The latter phenotype lacks gross morphology, glandular epithelium and gastric acid secretion has been described in independent teleost lineages and monotremes. Changes in stomach morphology have been correlated with the loss of gastric enzymes (*Pga*, *Pgc*), proton pump subunits *(Atp4a*, *Atp4b*) and the gastrin hormone (*Gast*) in monotremes (Ordoñez et al., 2008; Castro et al., 2014; Zhou et al., 2021, Kato et al., 2024). Due to the complex constitution and function of the stomach – with a large complement of cell types, signalling pathways and physiological functions – we hypothesised that the gene losses associated with the agastric phenotype extended beyond these aforementioned losses. To test this, we investigated the evolution, selection and loss of the gastrin physiological pathway along with genes implicated in gastric development, and functionality in both gastric and agastric Sarcopterygii and Actinopterygii taxa.

Gastrin physiological pathway: Correlations with gastric phenotype and genetic dispensability

We determined loss of gastrin in four of the five agastric species presented (platypus, echidna, pufferfish, West-African Lungfish) but surprisingly no substantial differences in substitution rate or selection pressures of the gastrin physiological pathway constituents (*Gast*, *Cckbr*, *Grp*, *Grpr*) between gastric and agastric species, or between species who had

retained or lost gastrin. This conservation is likely due to diversification in functions and signalling pathways beyond gastric acid secretion (Li et al., 2003; Albalat & Cañestro, 2016).

The release of gastrin occurs via gastric distension or gastrin-releasing peptide receptor activation and potentiates the release of gastric acid (Ordoñez et al., 2008). Despite retaining the receptor, the zebrafish (*D. rerio*) has lost the genes necessary to release gastric acid (*Atp4a*, *Atp4b*) and most gastric morphology (Castro et al, 2014). In contrast to other agastric fish species – zebrafish (*D. rerio*) has retained gastrin. Recent research has found novel roles of gastrin and its related isoforms (progastrin and the glycine-extended gastrins) in cell proliferation, apoptosis and extracellular remodelling in colonic epithelium in a suite of vertebrate models (Dimaline & Varro, 2014). Furthermore, mouse gastrin knockout models present with heavily deleterious gastric phenotypes: achlorydic gastric environments with high incidences of gastric metaplasia, bacterial overgrowth, and tumours (Friis-Hansen, 2007). A low substitution rate for zebrafish *Gast* or the native receptor *Cckbr*, indicates that gastrin may have further diversified in functionality (perhaps due to the *Cckbr* duplication) since the third whole-genome duplication event (3R) in teleosts, and maintenance of these functions explains the sequence conservation (Brunet et al., 2006; Ozernyuk & Schepetov, 2022). This also suggests that loss of gastrin in other agastric species may not only be due to loss of gastric epithelium and gastric acid secretion.

Further, we found no association between gastric phenotype (gastric vs agastric) or loss of gastrin with the absence of *Grpr* in the agastric platypus and gastric three-spined stickleback. The gastrin-releasing peptide pathway is implicated not only in gastrin release, but also in food intake regulation and growth in fish models, and in sexual arousal, thermoregulation and circadian rhythm regulation in rodent models (Volkoff et al., 2005; Duan, Rico & Merchant,

2022; Oti & Sakamoto, 2023; Zhao et al., 2023). The loss of GRP-GRPR signaling despite a retention of gastrin the three-spined stickleback suggests either that gastric mechanical stimuli may be sufficient to potentiate gastrin release in this species or that an independent gastrin-release mechanism is yet to be explored. This independent mechanism may be partially explained by the fact that the GRP and GRPR peptides are from the “bombesin-like” gene family, a family of three homologous lineages which diverged before the advent of jawed vertebrates and all with various roles as both neuropeptides and gastrointestinal hormones. Though affinity is weak, there is agonism between GRP and the homologous receptors.(Baldwin, Patel & Shulkes, 2007; Jensen et al., 2008).

Evolution of gastric genes implicated in the secondary loss of gastric phenotype in jawed vertebrates (Gnathostomata)

We noted the absence of gastrokine (*Gkn1*, *Gkn2*) and trefoil factor homologs (*Tff1*, *Tff2*), *Vsig1* and *Anxa10* in all agastric taxa to the exception of gastrokine-like sequences in the West-African lungfish. Convergent loss between agastric Actinopterygii taxa, the West-African lungfish and the platypus suggests a role in the secondary loss of stomach features in line with loss of gastric enzymes, proton pump subunits and gastrin (Ordoñez et al., 2008, Castro et al., 2014; Zhou et al., 2021; Kato et al., 2024). However, the echidna has an agastric phenotype and has retained *Gkn1*, *Tff2*, *Vsig1* and *Anxa10*. These findings align with our recent works in which the common monotreme ancestor likely lost antral glandular epithelium and pyloric restriction through *Nkx3.2* pseudogenisation and subsequent shift of developmental processes, but the echidna lineage displayed retention or re-establishment of pyloric-like restriction through potential compensation or an additional evolutionary event (Dann et al., 2024).

The presence of trefoil factor and gastrokine homologs in Actinopterygii taxa suggests the evolution of these genes around the divergence of jawed vertebrates (Gnathostomata), (Jiang & Applegate, 2011; Geahlen et al., 2013). Furthermore, presence of a homologous trefoil factor-like sequence in the vase tunicate (*Ciona intestinalis*) provides evidence of gene evolution in line with the advent of the alimentary canal in early ascidians (Table S1), (Nakayama & Ogasawara, 2017). Recent research has implicated these genes in the differentiation of squamous and glandular epithelium in the mammalian hindstomach as well as mucosal proliferation/differentiation and protection to chemical insult. The gastrokines and trefoil factor homologs, in particular, are known for their potent anti-tumorigenic properties and are implicated in several gastrointestinal cancers (Yan et al., 2011; Jahan et al., 2020). Further *in-vivo* experimentation in a variety of biological models could determine whether these are newly-acquired functions or whether these features date back to early jawed-vertebrates (Oidovsambuu et al., 2011, Gerke et al., 2005; Tsai et al., 2015).

Conclusion

Here, we investigated the evolution, selection and conservation of the gastrin physiological pathway and other gene involved in gastric function in gastric and agastric jawed vertebrates in order to further understand the independent evolutionary trajectories of the agastric phenotype. Interestingly, we discovered that the absence of a gastric gene repertoire (*Gkn1*, *Gkn2*, *Tff1*, *Tff2*, *Vsig1*, *Anxa10*) – which evolved prior to jawed vertebrate divergence – correlates with the agastric phenotype to the exception of the echidna, which may have retained some aspects of gastric functionality (Figure 4). We found that the loss of *Gast* and *Grpr* does not correlate with the gastric or agastric phenotype. Furthermore, we did not observe a shift in sequence evolution or selection associated with the loss of these genes in vertebrates. These findings differentiate aspects of the agastric phenotype and reveal convergent gene loss and retention in agastric species separated by more than 500 million years of evolution.

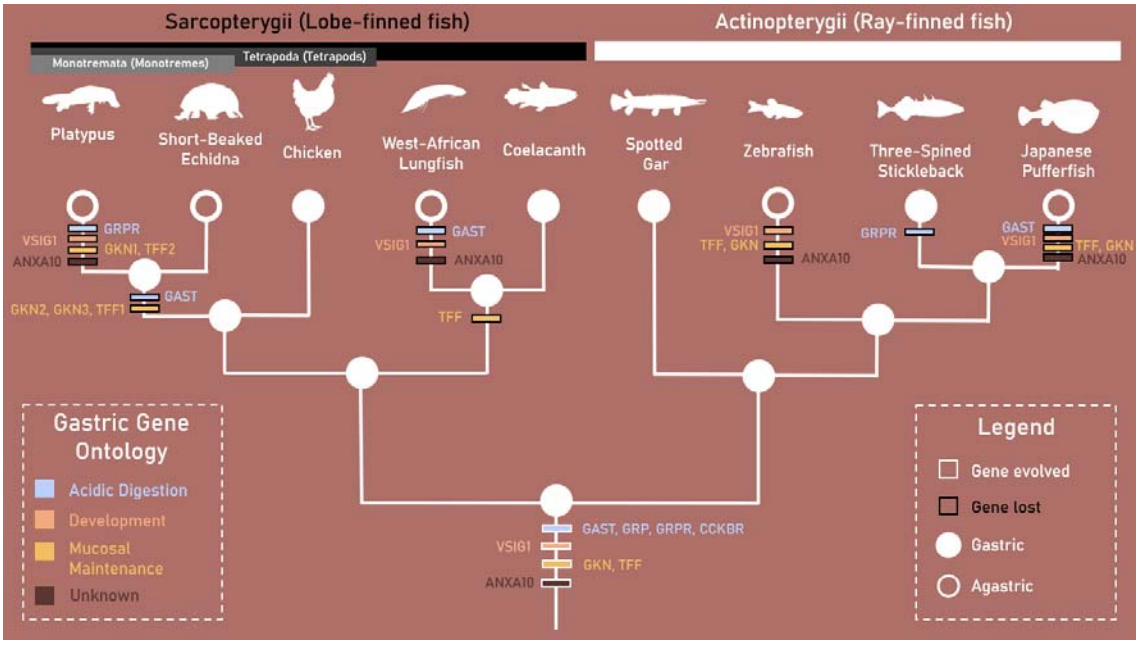


Figure 4: Phylogeny outlining the proposed gastric gene evolution, duplication and loss events between taxa displaying the gastric and agastric phenotype in Sarcopterygii (lobe-finned fish) and Actinopterygii (ray-finned fish) clades. Agastric taxa are denoted by hollow branch tips, gastric gene ontologies are outlined by gene colouration while gene evolution and loss events are signified by white and black borders respectively. Branches are not drawn to scale.

References

Agace, WW & McCoy, KD (2017), 'Regionalized Development and Maintenance of the Intestinal Adaptive Immune Landscape', *Immunity*, vol. 46, no. 4, pp. 532-548.

Albalat, R & Cañestro, C (2016), 'Evolution by gene loss', *Nature Reviews Genetics*, vol. 17, no. 7, pp. 379-391.

Apostopoulos, V, Stojanovska, L & Gargosky, SE (2015), 'MUC1 (CD227): a multi-tasked molecule', *Cellular and Molecular Life Sciences*, vol. 72, no. 23, pp. 4475-4500.

Baldwin, GS, Patel, O & Shulkes, A (2007), 'Phylogenetic analysis of the sequences of gastrin-releasing peptide and its receptors: Biological implications', *Regulatory Peptides*, vol. 143, no. 1, pp. 1-14.

Beasley, DE, Koltz, AM, Lambert, JE, Fierer, N & Dunn, RR (2015), 'The Evolution of Stomach Acidity and Its Relevance to the Human Microbiome', *PLoS One*, vol. 10, no. 7, p. e0134116.

Blomme, T, Vandepoele, K, De Bodt, S, Simillion, C, Maere, S & Van de Peer, Y (2006), 'The gain and loss of genes during 600 million years of vertebrate evolution', *Genome Biology*, vol. 7, no. 5, p. R43.

Blum, M, Chang, HY, Chugaransky, S, Grego, T, Kandasaamy, S, Mitchell, A, Nuka, G, Paysan-Lafosse, T, Qureshi, M, Raj, S, Richardson, L, Salazar, GA, Williams, L, Bork, P, Bridge, A, Gough, J, Haft, DH, Letunic, I, Marchler-Bauer, A, Mi, H, Natale, DA, Necci, M, Orengo, CA, Pandurangan, AP, Rivoire, C, Sigrist, CJA, Sillitoe, I, Thanki, N, Thomas, PD, Tosatto, SC, Wu, CH, Bateman, A & Finn, RD (2021), 'The InterPro protein families and domains database: 20 years on', *Nucleic Acids Research*, vol. 49, no. D1, pp. D344-D354.

Bolte, C, Ren, X, Tomley, T, Ustiyan, V, Pradhan, A, Hoggatt, A, Kalin, TV, Herring, BP & Kalinichenko, VV (2015), 'Forkhead Box F2 Regulation of Platelet-derived Growth Factor and Myocardin/Serum Response Factor Signaling Is Essential for Intestinal Development', *Journal of Biological Chemistry*, vol. 290, no. 12, pp. 7563-7575.

Brunet, FG, Crollius, HR, Paris, M, Aury, JM, Gibert, P, Jaillon, O, Laudet, V & Robinson-Rechavi, M (2006), 'Gene Loss and Evolutionary Rates Following Whole-Genome Duplication in Teleost Fishes', *Molecular Biology and Evolution*, vol. 23, no. 9, pp. 1808-1816.

Bürglin, TR (2008), 'Evolution of *hedgehog* and *hedgehog*-related genes, their origin from Hog proteins in ancestral eukaryotes and discovery of a novel Hint motif', *BMC Genomics*, vol. 9, no. 1, p. 127.

Bürglin, TR (2008), 'The Hedgehog protein family', *Genome Biology*, vol. 9, no. 11, p. 241.

Carbon, S, Ireland, A, Mungall, CJ, Shu, S, Marshall, B, Lewis, S, AmiGO Hub & Web Presence Working Group (2009), 'AmiGO: online access to ontology and annotation data', *Bioinformatics*, vol. 25, no. 2, pp. 288-289.

Casewell, NR (2016), 'Venom Evolution: Gene Loss Shapes Phenotypic Plasticity', *Current Biology*, vol. 26, no. 18, pp. R849-851.

Castro, LFC, Gonçalves, O , Mazan, S, Tay, BH, Venkatesh, B & Wilson, JM (2014), 'Recurrent gene loss correlates with the evolution of stomach phenotypes in gnathostome history', *Proceedings of the Royal Society B*, vol. 281, no. 1775, p. 20132669.

Corfield, AP (2015), 'Mucins: A biologically relevant glycan barrier in mucosal protection', *Biochimica Et Biophysica Acta*, vol. 1850, no. 1, pp. 236-252.

Dann, J, Qu, Z, Shearwin-Whyatt, L, van der Ploeg, R & Grützner, F (2024), 'Pseudogenization of NK3 homeobox 2 (*Nkx3.2*) in monotremes provides insight into unique gastric anatomy and physiology', *Open Biology*, vol. 14, no. 7, p. 240071.

Dimaline, R & Varro, A (2014), 'Novel roles of gastrin', *The Journal of Physiology*, vol. 592, no. 14, pp. 2951-2958.

Dixon, J, Strugala, V, Griffin, SM, Welfare, MR, Dettmar, PW, Allen, A & Pearson, JP (2001), 'Esophageal mucin: an adherent mucus gel barrier is absent in the normal esophagus but present in columnar-lined Barret's esophagus', *The American Journal of Gastroenterology*, vol. 96, no. 9, pp. 2575-2583.

Duan, S, Rico, K & Merchant, JL (2022), 'Gastrin: From Physiology to Gastrointestinal Malignancies', *Function*, vol. 3, no. 1, p. zqab062.

Dupré, D & Tostivint, H (2014), 'Evolution of the gastrin-cholecystokinin gene family revealed by synteny analysis', *General and Comparative Endocrinology*, vol. 195, pp. 164-173.

Friis-Hansen, L (2007), 'Lessons from the gastrin knockout mice', *Regulatory Peptides*, vol. 139, no. 1, pp. 5-22.

Galura, GM, Chavez, LO, Robles, A & McCallum, R (2019), 'Gastroduodenal Injury: Role of Protective Factors', *Current Gastroenterology Reports*, vol. 21, no. 8, p. 34.

Gao, F, Chen, C, Arab, DA, Du, Z, He, Y & Ho, SYW (2019), 'EasyCodeML: A visual tool for analysis of selection using CodeML', *Ecology and Evolution*, vol. 9, no. 7, pp. 3891-3898.

Geahlen, JH, Lapid, C, Thorell, K, Nikolskiy, I, Huh, WJ, Oates, EL, Lennerz, JKM, Tian, X, Weis, VG, Khurana, SS, Lundin, SB, Templeton, AR & Mills, JC (2013), 'Evolution of the human gastrokine locus and confounding factors regarding the pseudogenicity of GKN3', *Physiological Genomics*, vol. 45, no. 15, pp. 667-683.

Gerke, V, Creutz, CE & Moss, SE (2005), 'Annexins: linking Ca^{2+} signalling to membrane dynamics', *Nature Reviews Molecular Cell Biology*, vol. 6, no. 6, pp. 449-461.

Guindon, S, Dufayard, JF, Lefort, V, Anisimova, M, Hordijk, W & Gascuel, O (2010), 'New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Performance of PhyML 3.0', *Systematic Biology*, vol. 59, no. 3, pp. 307-321.

Griffiths, M (1978), *The Biology of Monotremes*, New York: Academic Press, pp. 77-81.

He, C (2015), 'Investigating Function and Evolution of Genes and Proteins Involved in Metabolic Control in Mammals', PhD thesis, The University of Adelaide, Adelaide.

He, C, Enkhjargal, TA, Myers, MA, Forbes, BE & Grützner, F (2013), 'Changes in the ghrelin hormone pathway maybe part of an unusual gastric system in monotremes', *General and Comparative Endocrinology*, vol. 191, pp. 74-82.

He, C, Myers, MA, Forbes, BE & Grützner, F (2015),' Immunohistochemical analysis of pancreatic islets of platypus (*Ornithorhynchus anatinus*) and echidna (*Tachyglossus aculeatus ssp.*), *Journal of Anatomy*, vol. 226, no. 4, pp. 373-380.

Herring, BP, Hoggatt, AM, Gupta, A & Wo, JM (2018), 'Gastroparesis is associated with decreased FOXF1 and FOXF2 in humans, and loss of FOXF1 and FOXF2 results in gastroparesis in mice', *Neurogastroenterology & Motility*, vol. 31, no. 3, p. e13528.

Hoffmann, W (2015), 'TFF2, a MUC6-binding lectin stabilizing the gastric mucus barrier and more (Review), *International Journal of Oncology*, vol. 47, no. 3, pp. 806-816.

Hoffmann, W (2020), 'Trefoil factor family (TFF) peptides and their diverse molecular functions in mucus barrier protection and more: Changing the paradigm', *International Journal of Molecular Sciences*, vol. 21, no. 12, pp. 1-20.

Huangfu, D & Anderson, KV (2006), 'Signaling from Smo to Ci/Gli: conservation and divergence of Hedgehog pathways from *Drosophila* to vertebrates', *Development*, vol. 133, no. 1, pp. 3-14.

Jahan, R, Shah, A, Kisling, SG, Macha, MA, Thayer, S, Batra, SK & Sukhwinder, K (2020), 'Odyssey of trefoil factors in cancer: Diagnostic and therapeutic implications', *Biochimica et Biophysica Acta: Reviews on Cancer*, vol. 1873, no. 2, p. 188362.

Jensen, RT, Battey, JF, Spindel, ER & Benya, RV (2008), 'Mammalian Bombesin Receptors: Nomenclature, Distribution, Pharmacology, Signaling, and Functions in Normal and Disease States', *Pharmacological Reviews*, vol. 60, no. 1, pp. 1-42.

Jiang, Z, Lossie, AC & Applegate, TJ (2011), 'Evolution of Trefoil Factor(s): Genetic and Spatio-Temporal Expression of Trefoil Factor 2 in the Chicken (*Gallus Gallus Domesticus*)', *PloS one*, vol. 6, no. 7, p. e22691.

Jin, X, Deng, L, Li, H, Zhang, Z, He, Q, Yang, C, Jiang, H, Zhu, XQ & Peng, L (2011), 'Identification and characterization of a serine protease inhibitor with two trypsin inhibitor-like domains from the human hookworm *Ancylostoma duodenale*', *Parasitology Research*, vol. 108, no. 1, pp. 287-295.

Kageyama, R, Ohtsuka, T & Kobayashi, T (2007), 'The Hes gene family: repressors and oscillators that orchestrate embryogenesis', *Development*, vol. 134, no. 7, pp. 1243-1251.

Karasov, WH, Douglas, AE & Ronald, T (2013), 'Comparative Digestive Physiology', *Comprehensive Physiology*, pp. 741-783.

Kato, A, Pipil, S, Ota, C, Kusakabe, M, Watanabe, T, Nagashima, A, Chen, AP, Islam, Z, Hayashi, N, Wong, MKS, Komada, M, Romero, MF & Takei, Y (2024), 'Convergent gene losses and pseudogenizations in multiple lineages of stomachless fishes', *Communications Biology*, vol. 7, no. 1, p. 408.

Kim, TH & Shivdasani, RA (2016), 'Stomach development, stem cells and disease', *Development*, vol. 143, no. 4, pp. 554-565.

Koelz, HR (1992), 'Gastric Acid in Vertebrates', *Scandinavian Journal of Gastroenterology*, vol. 27, no. S193, pp. 2-6.

Krause, WJ & Leeson, CR (1974), 'The gastric mucosa of two monotremes: the duck-billed platypus and echidna', *The Journal of Morphology*, vol. 142, no. 1, pp. 285-299.

Kim, O, Yoon, JH, Choi, WS, Ashktorab, H, Smoot, DT, Nam, SW, Lee, JY & Park, WS (2013), 'GKN2 Contributes to the Homeostasis of Gastric Mucosa by Inhibiting GKN1 Activity', *Journal of Cellular Physiology*, vol. 229, no. 6, pp. 762-771.

Li, WH, Gu, Z, Steinmetz, LM, Gu, X, Scharfe, C & Davis, RW (2003), 'Role of duplicate genes in genetic robustness against null mutations', *Nature*, vol. 421, no. 6918, pp. 63-66.

Lu, S, Wang, J, Chitsaz, F, Derbyshire, MK, Geer, RC, Gonzales, NR, Gwadz, M, Hurwitz, DI, Marchler, GH, Song, JS, Thanki, N, Yamashita, RA, Yang, M, Zhang, D, Zheng, C, Lanczycki, CJ & Marchler-Bauer, A (2020), 'CDD/SPARCLE: the conserved domain database in 2020', *Nucleic Acids Research*, vol. 48, no. D1, pp. D265-D268.

Lu, SH, Chen, YL, Shun, CT, Lai, JN, Peng, SY, Lai, PL & Hsu, HC (2011), 'Expression and prognostic significance of gastric-specific annexin A10 in diffuse- and intestinal-type gastric carcinoma', *Journal of Gastroenterology and Hepatology*, vol. 26, no. 1, pp. 90-97.

Mendonça, AG, Alves, RJ & Pereira-Leal, JB (2011), 'Loss of Genetic Redundancy in Reductive Genome Evolution', *PloS Computational Biology*, vol. 7, no. 2, p. e1001082.

Menheniott, TR, Kurklu, B & Giraud, AS (2013), 'Gastrokines: stomach-specific proteins with putative homeostatic and tumor suppressor roles', *American Journal of Physiology: Gastrointestinal and Liver Physiology*, vol. 304, no. 2, pp. G109-G121.

McCracken, KW & Wells, JM (2017), 'Mechanisms of embryonic stomach development', *Seminars in Cell & Developmental Biology*, vol. 66, pp. 36-42.

Menting, JG, Gajewiak, J, MacRaild, CA, Chou, DHC, Disotuar, MM, Smith, NA, Miller, C, Erchegyi, J, Rivier, JE, Olivera, BM, Forbes, BE, Smith, BJ, Norton, RS, Safavi-Hemami, H & Lawrence, MC (2016), 'A minimized human insulin-receptor-binding motif revealed in a *Conus geographus* venom insulin', *Nature Structural & Molecular Biology*', vol. 23, no. 10, pp. 916-920.

Mikita, S, Torrents, D & Bork, P (2006), 'PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments', *Nucleic Acids Research*, vol. 34, pp. W609-W612.

Mohammed Khan, J & Ranganathan, S (2009), 'A multi-species comparative structural bioinformatics analysis of inherited mutations in α-D-Mannosidase reveals strong genotype-phenotype correlation', *BMC Genomics*, vol. 10, no. 3, p. S33.

Nakayama, S & Ogasawara, M (2017), 'Compartmentalized expression patterns of pancreatic- and gastric-related genes in the alimentary canal of the ascidian Ciona intestinalis: evolutionary insights into the functional regionality of the gastrointestinal tract in Olfactores', *Cell and Tissue Research*, vol. 370, no. 1, pp. 113-128.

Nakazawa, K, Yamazawa, T, Moriyama, Y, Ogura, Y, Kawai, N, Sasakura, Y & Saiga, H (2013), 'Formation of the digestive tract in *Ciona intestinalis* includes two distinct morphogenic processes between its anterior and posterior parts', *Developmental Dynamics*, vol. 242, no. 10, pp. 1172-1183.

Nyamai, DW & Tastan Bishop, U (2019), 'Aminoacyl tRNA synthetases as malarial drug targets: a comparative bioinformatics study', *Malaria Journal*, vol. 18, no. 1, pp. 1-27.

Ochinnikov, S, Kamisetty, H & Baker, D (2014), ‘Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information’, *eLife*, vol. 2014, no. 3, p. e02030.

Oidovsambuu, O, Nyamsuren, G, Liu, S, Goring, W, Engel, W, Adham, IM & Gires, O (2011), ‘Adhesion protein VSIG1 is required for the proper differentiation of glandular gastric epithelia’, *PloS one*, vol. 6, no. 10, p. e25908.

Ordoñez, GR, Hillier, LDW, Warren, WC, Grützner, F, López-Otín, C & Puente, XS (2008), ‘Loss of genes implicated in gastric function during platypus evolution’, *Genome Biology*, vol. 9, no. 5, p. R81.

Oti, T & Sakamoto, H (2023), ‘Neuropeptidergic control circuits in the spinal cord for male sexual behaviour: Oxytocin-gastrin-releasing peptide systems’, *Journal of Neuroendocrinology*, vol. 35, no. 9, p. e13324.

Ozernyuk, N & Schepetov, D (2022), ‘HOX-Gene Cluster Organization and Genome Duplication in Fishes and Mammals: Transcript Variant Distribution along the Anterior-Posterior Axis’, *International Journal of Molecular Sciences*, vol. 23, no. 17, p. 9990.

Pandey, A & Braun, EL (2020), ‘Protein evolution is structure dependent and non-homogenous across the tree of life’, *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*.

Parisi, G & Echave, J (2001), ‘Structural Constraints and Emergence of Sequence Patterns in Protein Evolution’, *Molecular Biology and Evolution*, vol. 18, no. 5, pp. 750-756.

Pavone, LM, Del Vecchio, P, Mallardo, P, Altieri, F, De Pasquale, V, Rea, Silviana, Martucci, NM, Di Stadio, CS, Pucci, P, Flagiello, A, Masullo, M, Arcari, P & Rippa, E (2013), ‘Structural characterization and biological properties of human gastrokine 1’, *Molecular bioSystems*, vol. 9, no. 3, pp. 412-421.

Pettersson, E, Lundeberg, J & Ahmadian, A (2009), ‘Generations of sequencing technologies’, *Genomics*, vol. 93, no. 2, pp. 105-111.

Pietrokovski, S (1994), ‘Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins’, *Protein Science*, vol. 3, no. 12, pp. 2340-2350.

Purohit, R, Peng, DS, Vielmas, E & Ondrus, AE (2020), ‘Dual roles of the sterol recognition region in Hedgehog protein modification’, *Communications Biology*, vol. 3, no. 1, p. 250.

Raina, D, Ahmad, R, Rajabi, H, Panchamoorthy, G, Kharbanda, S & Kufe, D (2012), 'Targeting cysteine-mediated dimerization of the MUC1-C oncoprotein in human cancer cells', *International Journal of Oncology*, vol. 40, no. 5, pp. 1643-1649.

Raina, D, Praveen, A, Lee, J, Bharti, A, McKnight, JC, Sharma, P, Kharbanda, S, Kufe, D & Srinivasula, SM (2015), 'Characterization of the MUC1-C cytoplasmic domain as a cancer target', *PloS one*, vol. 10, no. 8, p. e0135156.

Renteria, ME, Gandhi, NS, Vinuesa, P, Helmerhorst, E & Mancera, RL (2008), 'A Comparative Structural Bioinformatics Analysis of the Insulin Receptor Family Ectodomain Based on Phylogenetic Information', *PloS one*, vol. 3, no. 11, p. e3667.

Saxena, AK & Klimbacher, G (2018), 'Comparison of esophageal submucosal glands in experimental models for esophagus tissue engineering applications', *Esophagus: Official Journal of the Japanese Esophageal Society*, vol. 16, no. 1, pp. 77-84.

Schloss, JA, Gibbs, RA, Makhijani, VB & Marziali, A (2020), 'Cultivating DNA Sequencing Technology After the Human Genome Project', *Annual Review of Genomics and Human Genetics*, vol. 21, no. 1, pp. 117-138.

Schubert, ML & Peura, DA (2008), 'Control of Gastric Acid Secretion in Health and Disease', *Gastroenterology*, vol. 134, no. 7, pp. 1842-1860.

Sievers, F, Wilm, A, Dineen, D, Gibson, TJ, Karplus, K, Li, W, Lopez, R, McWilliam, H, Remmert, M, Söding, J, Thompson, JD & Higgins, DG (2011), 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, vol. 7, no. 1, p. 539.

Soutto, M, Belkhiri, A, Piazuelo, MB, Blanca, M, Schneider, BG, Peng, D, Jiang, A, Washington, MK, Kokoye, Y, Crowe, SE, Zaika, A, Correa, P, Peek Jr, RM & El-Rifai, W (2011), 'Loss of TFF1 is associated with activation of NF-κB-mediated inflammation and gastric neoplasia in mice and humans', *The Journal of Clinical Investigation*, vol. 121, no. 5, pp. 1753-1767.

The UniProt Consortium (2021), 'UniProt: the universal protein knowledgebase in 2021', *Nucleic Acids Research*, vol. 49, no. D1, pp. D480-D489.

Trifinopoulos, J, Nguyen, LT, von Haasler, A & Minh, BQ (2016), "W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis", *Nucleic Acids Research*, vol. 44, no. W1, pp. W232-W235.

Tsai, JH, Lin, YL, Cheng, YC, Chen, CC, Lin, LI, Tseng, LH, Cheng, ML, Liau, JY & Jeng, YM (2015), 'Aberrant expression of annexin A10 is closely related to gastric phenotype in serrated pathway to colorectal carcinoma', *Modern Pathology*, vol. 28, no. 2, pp. 268-278.

Volkoff, H, Canosa, LF, Unniappan, S, Cerdá-Reverter, JM, Bernier, NJ, Kelly, SP & Peter, RE (2005), 'Neuropeptides and the control of food intake in fish', *General and Comparative Endocrinology*, vol. 142, no. 1, pp. 3-19.

Willhelm, BT, Marguerat, S, Aligianni, S, Codlin, S, Watt, S & Bähler, J (2011), 'Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast', *Genome Biology*, vol. 12, no. 8, p. 82.

Wilson, JM & Castro, LFC (2010), *The multifunctional gut of fish*, Academic Press, London, UK.

Wray, GA, Hahn, MW, Abouheif, E, Balhoff, JP, Pizer, M, Rockman, MV & Romano, LA (2003), 'The Evolution of Transcriptional Regulation in Eukaryotes', *Molecular Biology and Evolution*, vol. 20, no. 9, pp. 1377-1419.

Yan, GR, Xu, SH, Tan, ZL, Yin, XF & He, QY (2011), 'Proteomics characterization of gastrokine 1-induced growth inhibition of gastric cancer cells', *Proteomics*, vol. 11, no. 18, pp. 3657-3664.

Yu, J, Blom, J, Glaeser, SP, Jaenicke, S, Juhre, T, Rupp, O, Schwengers, O, Spänig, S & Goesmann, A (2017), 'A review of bioinformatics platforms for comparative genomics.

Recent developments of the EDGAR 2.0 platform and its utility for taxonomic and phylogenetic studies', *Journal of Biotechnology*, vol. 261, no. 10, pp. 2-9.

Yu, D, Ren, Y, Uesaka, M, Beavan, AJS, Muffato, M, Shen, J, Li, Y, Sato, I, Wan, W, Clark, JW, Keating, JN, Carlisle, EM, Dearden, RP, Giles, S, Randle, E, Sansom, RS, Feuda, R, Fleming, JF, Sugahara, F, Cummins, C, Patricio, M, Akanni, W, D'Aniello, S, Bertolucci, C, Irie, N, Cantas, A, Sheng, G, de Mendoza, A, Maeso, I, Irimia, M, Fromm, B, Peterson, KJ, Das, S, Hirano, M, Rast, JP, Cooper, MD, Paps, J, Pisani, D, Kuratani, S, Martin, FJ, Wang, W, Donoghue, PCJ, Zhang, YE & Pascual-Anaya, J (2024), 'Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences', *Nature Ecology & Evolution*, vol. 8, no. 3, pp. 519-535.

Zavros, Y, Waghray, M, Tessier, A, Bai, L, Todisco, A, Gumucio, DL, Samuelson, LC, Dlugosz, A & Merchant, JL (2007), 'Reduced Pepsin A Processing of Sonic Hedgehog in Parietal Cells Precedes Gastric Atrophy and Transformation', *Journal of Biological Chemistry*, vol. 282, no. 46, pp. 33265-33274.

Zhao, T Chen, A, Dai, D, Li, Z, Gao, XF & Xiong, L (2023), 'Role of the GRP/GRPR System in Regulating Brain Functions', *ACS Chemical Neuroscience*, vol. 14, no. 19, pp. 3588-3598.

Znalesniak, EB, Salm, F & Hoffmann, W (2020), ‘Molecular Alterations in the Stomach of *Tff1*-Deficient Mice: Early Steps in Antral Carcinogenesis’, *International Journal of Molecular Sciences*, vol. 21, no. 2, p. 644.