Recurrent gene loss correlates with the evolution of stomach phenotypes in gnathostome history

L. Filipe C. Castro¹,², Odete Gonçalves¹,², Sylvie Mazan³, Boon-Hui Tay⁴, Byrappa Venkatesh⁴ and Jonathan M. Wilson¹

¹CIMAR Associate Laboratory, Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), and ²Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal
³Development and Evolution of Vertebrates, CNRS-UPMC-UMR 7150, Station Biologique, Roscoff, France
⁴Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR (Agency for Science, Technology and Research), Biopolis, Singapore

The stomach, a hallmark of gnathostome evolution, represents a unique anatomical innovation characterized by the presence of acid- and pepsin-secreting glands. However, the occurrence of these glands in gnathostome species is not universal; in the nineteenth century the French zoologist Cuvier first noted that some teleosts lacked a stomach. Strikingly, Holocephali (chimaeras), dipnooids (lungfish) and monotremes (egg-laying mammals) also lack acid secretion and a gastric cellular phenotype. Here, we test the hypothesis that loss of the gastric phenotype is correlated with the loss of key gastric genes. We investigated species from all the main gnathostome lineages and show the specific contribution of gene loss to the widespread distribution of the agastric condition. We establish that the stomach loss correlates with the persistent and complete absence of the gastric function gene kit—H⁺/K⁺-ATPase (Atp4A and Atp4B) and pepsinogens (Pga, Pgc, Cym)—in the analysed species. We also find that in gastric species the pepsinogen gene complement varies significantly (e.g. two to four in teleosts and tens in some mammals) with multiple events of pseudogenization identified in various lineages. We propose that relaxation of purifying selection in pepsinogen genes and possibly proton pump genes in response to dietary changes led to the numerous independent events of stomach loss in gnathostome history. Significantly, the absence of the gastric genes predicts that reinvention of the stomach in agastric lineages would be highly improbable, in line with Dollo’s principle.

1. Introduction

Gene duplication is a powerful event underscoring evolutionary change [1]. In effect, genome duplications in vertebrate history (2R and 3R) have been linked with the evolution of novel morphological and physiological traits [2–5]. In parallel, the impact of gene loss on phenotypic variation and diversity has been less explored, though gene extinction events are frequent in evolution and have phenotypic impacts [6–9].

The stomach, a specialized segment of the gut, is characterized by the occurrence of acid (HCl) and pepsin-producing gastric glands [10]. It represents a notable case of morphological variability, as first noted by Aristotle (345 BC) while commenting in his Historia Animalium about ‘the diversity of shapes’ [11]. Gastric glands first appeared approximately 450 Myr ago and represent a key functional innovation found exclusively in jawed vertebrates (figure 1) [10]. Invertebrate chordates such as amphioxus and ascidians lack a stomach, as do the jawless lampreys and hagfishes. Evolutionarily, the acquisition of an acidic luminal environment extended the spectrum of dietary protein sources. At low pH, long proteins are denatured, facilitating the action of endopeptidases, which also evolved to operate optimally at these low pH levels [10,12]. The improvement by low pH of phosphate [13] and calcium [14,15] uptake, as well as its effect as a barrier against pathogen entry to the intestine, were also probably beneficial [10].
From the eighteenth to the twentieth century, a string of findings revealed the significance of gastric acid secretion in digestion, as well as the identification of its functional components, the pepsin and the gastric proton pump [16]. The gastric proton pump is an H^+/K^+-ATPase belonging to the P-type IIc ATPase subfamily that is capable of pumping H^+ against a 160 mM gradient [17]. It is a heterodimer composed of unrelated subunits: the HK\alpha (alpha subunit), encoded by \textit{Atp4A}, and the HK\beta (beta subunit), encoded by \textit{Atp4B}. \textit{Atp4A} has been isolated and characterized in various vertebrate classes, including the Atlantic stingray, a cartilaginous fish [18,19]. The gastric \textit{Atp4B} is unique among the P-type IIc family of ATPase beta subunit genes because it is the sole part of the gastrin-releasing peptide receptor, which is activated at low pH and digest proteins into smaller peptides, making them available for further digestion and absorption [12]. The diversity of pepsinogen gene families has been suggested to underscore differences in substrate specificity [12,21].

Surprisingly, acid–pepsinogen-producing glands are absent in several gnathostome lineages. The French zoologist Georges Cuvier (1805) [22], on the basis of gross morphology, first observed that some teleost groups (e.g. cyprinidae and labridae) lack a stomach. The number of observations of the agastic phenotype in teleost fishes has expanded greatly over the past 200 years. A conservative estimate indicates the occurrence of 15 independent loss events covering seven families and 20–27% of species [23]. Significantly, the lack of gastric acid secretion is not limited to the teleost lineage; Holoccephali, dipnoids and monotremes also exhibit an agastic gut (figure 1) [24]. The molecular foundations for the occurrence of distinct stomach phenotypes in gnathostome lineages are still unclear. The lack of gastric glands in platypus coincides with the loss of the gene repertoire involved in gastric digestion [25]. Here, we test the hypothesis that absence of the gastric phenotype is correlated with absence of key gastric genes. We investigate whether the extensive distribution of the agastic phenotype in vertebrate history is paralleled by differences in gene complement, namely the pepsinogens and the \textit{Atp4A}/\textit{Atp4B} genes. Taking advantage of full genome sequences from all major vertebrate classes, we show that the lack of a stomach phenotype (acid secretion and pepsin activity) correlates with the targeted deletion or inactivation of the gene repertoire involved in the gastric function, in all of the examined species.

2. Material and methods
(a) Database searches and synteny analysis
The full coding sequences of the human proteins for \textit{ATP4A}, \textit{ATP4B} and pepsinogen genes were used as query for TBLASTN searches of the Ensembl (release 69) and GenBank databases. The genomes of 14 vertebrate species were investigated: \textit{Homo sapiens} (human, gastric, GRCh37), \textit{Mus musculus} (mouse, gastric, GRCm38), \textit{Monodelphis domestica} (opossum, gastric, BROADOS), \textit{Gallus gallus} (chicken, gastric, WASHUC2), \textit{Anolis carolinensis} (anolis, gastric, AnoCar2.0), \textit{Xenopus tropicalis} (western clawed frog, gastric, JGI_4.2), \textit{Gasterosteus aculeatus} (stickleback, gastric, BROADS1), \textit{Oryzias latipes} (medaka, agastic, MEDAKA1), \textit{Takifugu rubripes} (pufferfish, agastic, FUGA), \textit{Tetraodon nigroviridis} (green-spotted pufferfish, agastic, TETRAODON8), \textit{Gadus morhua} (cod, gastric, gadMor1), \textit{Oreochromis niloticus} (Nile tilapia, gastric, Onril1.0), \textit{Xiphophorus maculatus} (platyfish, agastic, Xipmac v. 4.4.2) and \textit{Danio rerio} (zebrafish, gastric, Zv. 9). The genomic location of each gene was verified through phylogenetics (not shown). The \textit{Callorhinus milii} genome was searched with TBLASTN at http://esharkgenome.imcb.a-star.edu.sg/Blast/.

(b) Alignment and phylogenetics
The retrieved protein sequences were aligned using MAFFT with the L-INS-i method [26]. The GenBank accession numbers or Ensembl codes for each sequence are indicated in electronic supplementary material S1. Gaps were removed from each alignment. The full coding sequences of the human proteins for \textit{Atp4A} in \textit{vivo} [17,20]. Fundamental to gastric function are the pepsinogens, such as Pepsinogen \textit{A} (Pga), Pepsinogen \textit{B} (Pgb), Pepsinogen \textit{F} (Pgf), Progastricacin (Pgc) and Prochymosin (Cym) [12]. These are precursors of the pepsin enzymes, which are activated at low pH and digest proteins into smaller peptides, making them available for further digestion and absorption [12]. The diversity of pepsinogen gene families has been suggested to underscore differences in substrate specificity [12,21].

Figure 1. Phylogenetic relationships of extant chordate lineages. Plus and minus signs represent the presence or the absence of gastric glands, respectively. Timings of whole genome duplications (2R and 3R) are indicated. (Online version in colour.)
were rooted using *Ciona intestinalis* as an outgroup. The pepsinogen tree was rooted with CTSE-like sequences.

(c) Isolation of chicken *Atp4A* and catshark *Atp4B*

Chicken (*G. gallus*) specimens were obtained commercially in Portugal and killed by cervical dislocation prior to tissue sampling. Catsharks (*Squaliornis canicula*) were obtained locally and anaesthetized with tricaine methanesulfonate (1:5000 MS-222 Aquapharm UK) pH 7.5 and killed by cervical transection. Animals were treated in accordance with the Portuguese Animal Welfare Law (Decreto-Lei no.197/96) and animal protocols approved by CIIIMAR/UP and DGV (Ministry of Agriculture, UK). RNA was isolated from gastric samples with Illustra RNAspin mini kits with column DNase I treatment (GE Healthcare, UK). Concentration and purity of RNA were assessed and 1 μg of total RNA converted into first strand cDNA with oligo-dT primer and *SuperScript III Reverse transcriptase* (Life Technologies, USA). The polymerase chain reaction (PCR) and rapid amplification of cDNA ends (RACE) methods (SMARTer RACE, Clontech, USA) were used to isolate complete ORF of *Atp4A* and *Atp4B* in chicken and catshark. The Phusion Flash hot start high fidelity polymerase mix was used in the vicinity of the teleost *Atp4A* represented here by stickleback, cod and tilapia (figure 3b). In the agastric teleost species, even though a strongly conserved synteny block was found, there was no evidence for an intervening *Atp4A* gene ORF. Surprisingly, we were unable to find in the current chicken genome assembly hits similar to *Atp4A*, or any of the flanking gene families (e.g. *Tmco3*). The finding was surprising since birds undoubtedly have HCl secretion was unsequenced in the current chicken genome we undertook a PCR approach to isolate a complete *Atp4A* sequence. Our results indicate that the gene is present in the chicken genome (figure 2a). Independently, we also found *Atp4A* orthologue in two other birds: *Falcoperegrinus* (peregrine falcon) and *Pseudopodoces humilis* (Tibetan ground-tit).

Regarding the *Atp4B* loci, a strongly conserved gene arrangement was found in both tetrapods and teleosteans (figure 3b). Again, the agastric species have no intervening *Atp4B* ORF. Thus, we find an absolute correlation between the agastric phenotype and the absence of the two proton pump genes in their expected genomic loci. The preservation of the gene loci arrangements in combination with the phylogenetic analysis indicates that gene deletion has taken place in the genome of all four agastric teleosteans, paralleling similar findings in the stomach-less mammal, the platypus [25].

(b) Duplication, pseudogenization and loss of pepsinogen genes in vertebrate species

Tetrapod aspartic proteases expressed almost uniquely in the stomach include *Pga*, *Pgf*, *Pgc*, *Pgb* and *Cym* [12,32–34]. In fishes, a distinct nomenclature is currently used and three pepsinogen gene families have been described: *Pgc*, *Pga1* and *Pga2* [35–38].

We began by determining the repertoire of *Pga*, *Pgf* and *Cym* in tetrapods. *Pgc* and *Pgb* were not examined here since we previously documented their history in tetrapods [38]. Both phylogenetic methods recovered similar tree topologies although with some differences (see electronic supplementary material S3). For example, *Cym* sequences form a monophyletic assembly in the NJ tree, while in the ML tree the mammalian and sauropsid sequences are separated. In general, higher bootstrap support was found in the NJ analysis, although the more basal nodes display lower values (see electronic supplementary material, S3). A single *Cym* gene was found in all of the examined tetrapod species, except in the western clawed frog and human. In the latter, a *Cym* pseudogene has been previously reported [39]. The number of *Pga* genes varies significantly between the studied species. Three clear paralogues are found in humans, although they represent an independent duplication. In chicken, we found two *Pga*-like genes, although the sequence of the second gene was very short and was not included in the phylogeny (not shown). Single-copy *Pga* orthologues were recovered in anolis and the western clawed frog. In contrast to previous suggestions [25],
we found no indication for $P_{ga}$ or $P_{gf}$ in the opossum. In the analysed species, we only identified $P_{gf}$ in mammalian species. One sequence from the western clawed frog groups with mammalian $P_{gf}$, although with low statistical support to infer a proper orthology. This finding suggests that the $P_{ga}/P_{gf}$ duplication pre-dates tetrapod radiation.

We next inspected the gene loci of $P_{ga}$, $P_{gf}$ and $Cym$ (figure 4). A strong conservation in gene arrangement was detected, with $P_{ga}$ and $P_{gf}$ residing in the same locus. In the opossum, we were able to determine the existence of two pseudogenes with similarity to $P_{ga}/P_{gf}$, which explains the conflicting results with Ordoñez et al. [25] (figure 4;
electronic supplementary material S4a). The combination of phylogenetics and the analysis of the genomic loci supports the interpretation that Pgf was lost at least in humans and opossum, and probably in anolis and chicken.

In the analysed teleost species, a substantial degree of variation was found with respect to the pepsinogen gene complement. The highest gene number was found in stickleback and cod, each with four pepsinogen genes. In tilapia, three
pepsinogen-like sequences were recovered, while a single gene was found in pufferfishes (Takifugu and Tetraodon). In zebrasfish, platyfish and medaka no relevant sequences with similarity to pepsinogen genes were discovered. The phylogenetic analysis provided clarification to the orthology of the retrieved sequences (see electronic supplementary material S3). Three of the sequences found in the stickleback and cod correspond to orthologues of Pgc, Pga1 and Pga2. The fourth sequence represents a novel Pga-like gene clade not previously reported, which we name Pga3. In tilapia, one sequence robustly groups with Pga2 clade, while the remaining groups with Pga1. The analysis of genomic loci of the various pepsinogen genes shows that these are mostly conserved between teleost species (figure 5). Surprisingly, even though teleost Pga genes form an independent clade, the gene families in the vicinity of Pga2 and Pga3 suggest that these are orthologues of tetrapod Cym and not Pga as previously reported [32]. Finally, the conservation of the loci in combination with the phylogenetic analysis provides support that extensive gene loss has taken place. With the exception of pufferfishes, no pepsinogen genes were found in the other agastic teleost lineages.

Previously, we also established that Pgc orthologues are absent in agastic teleosts such as medaka, zebrasfish, Takifugu and Tetraodon, while present in the gastric stickleback [38]. Here, we extend the analysis to include two other teleost species: cod and tilapia (see electronic supplementary material S4b). We find that Pgc is present in cod but absent in tilapia. The lack of Pgc in tilapia is noticeable, as this species has a clear gastric phenotype [40], supported by our finding of Atp4A and Atp4B orthologues. Examination of the 'Pgc' locus in this species provides strong evidence for pseudogenization.

In summary, we find that the stomach phenotype (presence/absence of gastric glands) positively correlates with the complement of pepsinogen genes in the vertebrate species, although we also find gene retention in agastic species (e.g. Takifugu) and gene loss in gastric ones (e.g. tilapia).

(c) Atp4b is missing in the stomach-less chimaera Callorhinichus milii, but present in the gastric elasmobranch Scyliorhinus canicula

We next investigated the earliest branch of jawed vertebrates, the chondrichthians (cartilaginous fishes), where gastric and agastic lineages have been documented (agastic Holoccephali [41] and gastric Elasmobranchii [42]; figure 1). We started by searching the genome sequence of the Holoccephali shark C. milii [43]. No relevant hits were found for the Atp4A and Atp4B genes as well as pepsinogens (not shown). Nevertheless, the low coverage of the genome (1.4×) and the absence of synteny data impedes a firm conclusion. Thus, we decided to determine whether true gene loss justified our negative findings. We selected the Atp4B gene locus for analysis given the striking conservation of the flanking gene families among the analysed species. Typically, Grk1 and Tmco3 flank Atp4B in both teleosts and tetrapods (figure 3). We identified orthologues of these genes in the C. milii genome and isolated two BAC clones (2S121 and 132E7) that contained both genes. Sequencing and analysis of the intergenic sequence showed no evidence for the presence of the Atp4B gene between C. milii Tmco3 and Grk1 genes (see electronic supplementary material S5a).

PCR with Atp4B degenerate primers using genomic C. milii
DNA was also negative (data not shown). Additionally, to support our prediction of secondary loss in the Holoccephali, we next attempted to isolate the Atp4B orthologue in the gastric S. canicula, a member of the sister group, the Elasmobranchii. Through a combination of PCR strategies, we isolated a gene sequence in S. canicula, which was confirmed to be an orthologue of Atp4B by phylogenetic analysis (figure 2b). Expression analysis shows that the isolated gene is expressed as predicted in the stomach, an indirect indication of its functionality (see electronic supplementary material S5b). Overall, the analysis of genes in fishes representing the two lineages of cartilaginous fishes, the Holoccephali and Elasmobranchii, provided further support for the hypothesis that the absence of stomach function genes (e.g. Atp4B) correlates with the lack of gastric glands.

4. Discussion

Stomach acid-peptic digestion first evolved in gnathostome ancestry [10]. Nevertheless, the early findings of Cuvier [22] suggested that gastric glands are not a general anatomical trait among vertebrate species. In fact, the agastric phenotype is surprisingly common in gnathostomes, and gastric glands are absent from numerous teleost fish species, Chimaeriformes (extant Holoccephali lineage), dipnoans and monotremes [23,25,41,44]. The distribution of both phenotypes and the phylogenetic relationships between extant vertebrate classes provides a clear indication that the loss of gastric glands represents a secondary event. Previously, it was shown that the agastric phenotype in the platypus correlates with the absence of the proton pump genes, Atp4A and Atp4B, and the pepsinogen genes [25]. Here, we investigated the repertoire of the gastric function genes in a diverse array of vertebrate species where gastric and agastric representatives are present. Using a combination of phylogenetics and comparative genomics, we show that the lack of stomach coincides with the recurrent targeted deletion of the proton pump genes (Atp4A and Atp4B) and the various pepsinogen gene families in all of the examined species. In the exceptional cases of the agastric Takifugu and platypus, pepsinogen genes have been retained only to serve non-digestive functions [25,35]. Although we have not investigated dipnoans, we expect that this lineage should also have undergone loss of gastric genes.

By providing a strong linkage between loss of the gastric phenotype and the loss of the functional genes for acid-peptic digestion, we give evidence for the irreversibility of stomach loss, consistent with Dollo’s principle.1 Such gene loss precludes the reinvention of complex traits [45]. This contrasts with a number of other studies demonstrating the reinvention of complex traits (e.g. wings in stick insect, coiling in snails, limb development in reptiles [46]). However, in these cases, gene loss and recovery have not been demonstrated but rather the genes involved are retained for functioning at different development stages or in other processes. There is no evidence for any alternative acid-excreting mechanism being involved in gastric acidification, and the gastric proton pump is highly conserved [10]. A missing link, lacking the gastric phenotype but possessing the gastric acid-peptic genes, and thus potentially capable of re-inventing the stomach, has yet to be identified.

Why exactly did the stomach undergo so many secondary episodes of loss during vertebrate evolution is a priori difficult to establish, as the conditions experienced by the ancestor may no longer be present in the extant species of that lineage. Presumably, there is a high cost in maintaining a complex organ like the stomach; however, estimates of these costs from specific dynamic action measurements remain ambiguous [23,47–49]. Nevertheless, acid pumping by the $H^+\mathord{/}K^+$-ATPase, protection of the stomach lining from the acid and neutralization of the acid in the intestine will all contribute [23,47]. By contrast, the presence of gastric glands represents a selective advantage by extending and facilitating the digestion of dietary proteins through the action of pepsins. These endopeptidases are capable of hydrolysing peptide bonds at low pH and act at the first stage of protein digestion. They probably have had their evolution affected by the dietary peptide composition [12,32,50]. Moreover, relative rate ratio tests have identified that each round of pepsinogen gene duplication is characterized by adaptive evolution [32]. Positive selection has been documented also in four
lysine residues found in the primate Pga specific gene lineage (designated Pgs2), with anticipated modifications of the catalytic activity [50]. In effect, the diversity of pepsinogens is reflected in their hydrolytic specificities against different proteins and peptides [12,30–52]. For example, the hydrolytic activity of human Pgc towards haemoglobin is 100% and almost 0% for the porcine Pgb [12]. The Pgb from dog is also unable to hydrolyse haemoglobin, whereas it efficiently hydrolyses gelatin [53]. The cleavage pattern of the pepsins from the rice field eel against a set of proteins (e.g. cytochrome c) is similar for the three isoforms (PG1, PG2 and PG3), but PG1 showed a higher cleavage efficiency [54]. Thus, protein dietary sources are prime selective drivers for functional pepsinogen diversification (positive selection). While new peptide components can lead to the retention and/or diversification of novel pepsinogen duplicates, the opposite is a possibility as well. As species colonize different environments, new peptide sources are integrated into the diet, while others disappear. Therefore, these alterations might render the activity of some pepsin enzymes ineffective or worthless, leading to the relaxation of negative selection. If so, mutational inactivation would accumulate in pepsinogen genes, leading to gene loss. We consistently found numerous cases of ongoing pseudogenization in gastric species (electronic supplementary material S6). For example, CYM is inactive in humans [39] as a possible consequence of the lower level of proteins in human milk [33], although it is suggested to be essential for neonatal milk digestion in mammals [55]. Pga and Pgb in the opossum, Pga and Pgb in the mouse, Pgc in tilapia, and a copy of Pgb in anolis are also undergoing processes of pseudogenization [38] (this study; electronic supplementary material S6). Given the complexity of dietary protein components no clear pattern emerges as to what exact component triggers gene loss (electronic supplementary material S6), and this is probably variable between species. Nevertheless, in mammals pepsin levels (and gene numbers) in the stomach have been suggested to be related to the type of diet, with higher levels in herbivores [50]. Taken together, the overall findings suggest that pepsinogen genes are under distinct selective pressures in response to variable protein dietary components. Under this scenario, when the full complement of pepsinogen genes is lost due to hydrolytic activities no longer being required, the need of an acidic gastric environment disappears, causing the loss of gastric function (figure 6). An alternative scenario involves the loss of acidification of the stomach lumen with the pseudogenization of the pump genes that could secondarily trigger the complete loss of pepsinogen genes since an acid environment is fundamental for activation and activity. Dietary and/or environmental factors may again have created the relaxation of purifying selective pressures or emergence of neutral selective pressures for loss of gastric acidification. Possible scenarios include a switch to a diet rich in calcareous organisms (high buffer capacity; e.g. coral feeding parrot fish Scaridae) or feeding on large amounts of barely digestible material (sediment and detritus, e.g. many Cyprinidae) that neutralize gastric acidification, rendering the pump genes superfluous [25,56].

In conclusion, the findings reported clearly illustrate a remarkable case of the role of gene loss in phenotypic variability. We find that the simplification of the vertebrate gut, with the loss of gastric glands, has occurred multiple times in vertebrate evolution as a result of the loss of genes central to acid-peptic digestion.

Acknowledgements. We thank two anonymous reviewers for improving a previous version of the manuscript.

Funding statement. Work supported by Portuguese Foundation for Science and Technology (grant no. PTDC/MAR/98035/2008 and SFRH/BD/79821/2011). This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE—Operational Competitiveness Programme and national funds through FCT—Foundation for Science and Technology, under the project PEst-C/MAR/LA0015/2013. S.M. was supported by Région Bretagne (grant no. 079753 EVOVERT). Work in B.V.’s lab was supported by the Biomedical Research Council of A*STAR, Singapore.

Endnote

1The idea that complex characters that have been lost in evolution cannot be regained [45].

References


