Horizontal transfer and gene loss shaped the evolution of alpha-amylases in bilaterians

Andrea Desiderato (1, 2), Marcos Barbeitos (1), Clément Gilbert (3), Jean-Luc Da Lage (3)

(1) Graduate Program in Zoology, Zoology Department, Federal University of Paraná, CP 19020, Curitiba, Paraná 81531-980, Brazil

(2) Department of Functional Ecology, Alfred Wegener Institute & Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

(3) Évolution, Génomes, Comportement, Écologie. CNRS, IRD, Université Paris-Sud. Université Paris-Saclay. F-91198 Gif-sur-Yvette, France

Abstract

The subfamily GH13_1 of alpha-amylases is typical of Fungi, but it also includes some unicellular eukaryotes (e.g. Amoebozoa, choanoflagellates) and non-bilaterian Metazoa. Conversely, since a previous study in 2007, all Bilateria were considered to harbor only alpha-amylases supposedly inherited by horizontal transfer from a proteobacterium and classified in the subfamilies GH13_15 and 24, which were therefore commonly called bilaterian alpha-amylases. The taxonomic scope of Eukaryota genomes in databases has been greatly increased ever since 2007. We have surveyed GH13_1 sequences in recent data from non-bilaterian animals and unicellular eukaryotes. We found a number of those sequences in Anthozoa

(Cnidaria) and in sponges, confirming the previous observations, but none in Ctenophora. Most surprisingly, such fungal (also called Dictyo-type) amylases were also consistently retrieved in a limited number of bilaterian phyla: hemichordates (deuterostomes), brachiopods, some molluscs and annelids (protostomes). We discuss evolutionary hypotheses for these findings, namely, the retention of the ancestral gene in those phyla only and/or horizontal transfers from non-bilaterian donors.

Key words: alpha-amylase, gene loss, horizontal gene transfer, hemichordates, brachiopods, phoronids, molluscs, annelids, Bilateria, glycosyl hydrolase, introns

Introduction

Alpha-amylases are enzymes that are almost ubiquitous in the living world, where they perform the hydrolysis of starch and related polysaccharides into smaller molecules, to supply energy to the organism through digestion. They belong to glycosyl hydrolases, a very large group of enzymes which have been classified in a number of families according to their structures, sequences, catalytic activities and catalytic mechanisms (HENRISSAT AND DAVIES 1997). Most alpha-amylases are members of the glycoside hydrolase family 13 (GH13), which includes enzymes that can either break down or synthetize α -1,4-, α -1,6- and, less commonly, α -1,2- and α -1,3-glycosidic linkages. Sucrose and trehalose are also substrates for enzymes of this family (MACGREGOR *et al.* 2001). The numerous family GH13 is divided into 42 subfamilies, of which only three occur in Metazoans: GH13_1, GH13_15 and GH13_24 (STAM *et al.* 2006; DA LAGE *et al.* 2007)*;* Lombard, 2014 #2103}. The latter two include the common animal alpha-amylases, while the former was first described in Fungi for which it represents the canonical alpha-amylase (STAM *et al.* 2006). In 2007, Da Lage et al. (DA LAGE *et al.* 2007) described the subfamilies GH13_15/24 as private to Bilateria among metazoans. In the same article, they retrieved sequences belonging to the subfamily GH13_1 from the sponge *Amphimedon queenslandica* (named *Reniera sp*. in their paper) and the sea anemone *Nematostella vectensis*, besides the unikont choanoflagellates and amoebozoans, and also excavates and ciliates. They dubbed "Dictyo-type" this alpha-amylase, referring to the slime mold *Dictyostelium discoideum* (Amoebozoa Mycetozoa). The authors proposed that this amylase, ancestral to the Unikont clade, is shared among non-bilaterian metazoans (e.g. sponges, sea anemones and corals, and Placozoa), but is not found in the Bilateria, being replaced in this clade by an alpha-amylase of bacterial origin, whose sequence is close to the typical animal amylases.

Given that a number of new genomes have been sequenced in the twelve years after that publication, we explore again the diversification of this enzyme subfamily among the Eukaryota. We will focus mainly on Metazoa, in which we show unexpected situations of co-occurrence of both subfamilies GH13_1 and GH13_15 in the same genomes. We will discuss two mutually exclusive explanations that may be proposed: either the retention of the ancestral GH13_1 gene along with the typical bilaterian GH13_15 in multiple phyla, or horizontal transfer(s) from non-bilaterian donor(s) which would have to be identified.

Materials and methods

In order to further characterize the distribution of GH13_1 genes in Metazoa, we used the sequence of the sponge *Amphimedon queenslandica* GH13_1 (GenBank XP_019851448) as a query to perform BLASTP and TBLASTN searches on various online databases available in Genbank (nr, proteins, genomes, assembly, SRA, TSA, WGS), compagen.org, marinegenomics.oist.jp, reefgenomics.org, marimba.obs-vlfr.fr, vectorbase.org, AmpuBase (https://www.comp.hkbu.edu.hk/~db/AmpuBase/index.php) (IP *et al.* 2018), between October 2018 and April 2019. Fungi were not searched further in this study because they are known to have a GH13_1 member as the usual alpha-amylase. To increase the chances to retrieve potential cnidarian or ctenophoran sequences, the starlet sea anemone *Nematostella vectensis* amylase (XP_001629956) was also used to query those databases. After the discovery of GH13_1-like sequences in Bilateria, the sequence XP_013396432 of the brachiopod *Lingula anatina* was also used specifically for additional search in Bilateria. Non-animal eukaryote species were investigated using the *Dictyostelium discoideum* sequence XP_640516 as query. The BLAST hits were considered further when expectation values (e-values) were better (lower) than 10^{-100} for BLASTP or 10^{-75} for TBLASTN, except for constitutively small hits such as SRA (sequence read archives). These were only considered when several highly significant hits covered most of the query sequence. When SRA hits had too many gaps, we did not attempt to assemble longer sequences and thus we did not use such sequences in alignments or phylogenies. Finally, we kept only sequences which were inside long contigs, or full-size or near full-size transcripts. We also checked once again the absence of animal-type alpha-amylase (GH13_15 or 24) outside the Bilateria using the sequence of the bivalve *Corbicula fluminea* (AAO17927) as a BLASTP query. The CAZy database (cazy.org (LOMBARD *et al.* 2014)), which is devoted to glycosyl hydrolases and related enzymes was used to check assignment of some of the sequences found to the GH13_1 subfamily.

Intron-exon gene structures were recovered either from alignments between genomic sequences and their mRNA counterparts, or using annotated graphic views when available in the databases. In some cases, for unannotated genes, the N-terminal and/or the C-terminal parts of the retrieved genomic sequences were uncertain, and were not retained in the analyses.

Alignments were performed using MUSCLE (EDGAR 2004), as implemented in Geneious (Biomatters Ltd.). A maximum likelihood (ML) tree was built using PhyML's (GUINDON AND GASCUEL 2003) current implementation at the phylogeny.fr portal (DEREEPER *et al.* 2008). To this end, we first trimmed the N-terminal protein sequences up to the first well

conserved LLTDR motif. C-terminal parts were also truncated at the last well aligned stretch. Gaps were removed from the alignments and data were analyzed under WAG (WHELAN AND GOLDMAN 2001) with among-site rate variation modeled by four discretized rate categories sampled from a gamma distribution. Both the alpha parameter and the proportion of invariable sites were estimated from the data. The robustness of the nodes was estimated using an approximate likelihood ratio test (aLRT) (ANISIMOVA AND GASCUEL 2006). The tree was drawn at the iTOL website (LETUNIC AND BORK 2016). Metazoans and choanoflagellates were clustered as the ingroup.

Results

The new sequences retrieved from the databases are listed in Table 1. A general protein alignment of the sequences found in this study along with already known GH13_1 sequences is shown in Fig. S1.

GH13_1 sequences retrieved from unicellular taxa

We confirmed the presence of GH13_1 in dictyostelids, in ciliates and also in oomycetes, some representatives of which (but not all) are indicated in Table 1. In two oomycetes, *Saprolegnia diclina* and *Achlya hypogyna*, the GH13_1-like sequences were the C-terminal half of longer sequences, the N-terminal half of which was similar to unclassified GH13 sequences found in e.g. *Acanthamoeba histolytica* (GenBank accession BAN39582), according to the CAZy database. In our general phylogenetic tree (Fig. 1), these sequences were used as outgroups. In choanoflagellates, where *Monosiga brevicollis* was already known to harbor a GH13_1 sequence (DA LAGE *et al.* 2007), we found a GH13_1 sequence in the genome of *Salpingoeca rosetta*. A partial sequence was also returned from incomplete genome data from *Monosiga ovata* (at Compagen, not shown).

GH13_1 sequences retrieved from non-bilaterian animals

In Cnidaria, a number of GH13_1 sequences were recovered from many Anthozoa species, (sea anemones, corals and allies), from genome as well as transcriptome data, at the

Reefgenomics database. Interestingly, we found no alpha-amylase sequences at all in Medusozoa (jellyfish, hydras) nor in Endocnidozoa (parasitic cnidarians). In the general tree (Fig. 1), cnidarian sequences form a clear cluster, with two main branches, grouping Actiniaria (sea anemones) and Pennatulacea (soft corals) on one branch, and Scleratinia (hard corals) and Corallimorpharia (mushroom anemones) on the other branch.

In sponges (Porifera), data were less abundant. No alpha-amylase sequence was found in *Sycon ciliatum* (Calcarea) and *Oscarella carmela* (Homoscleromorpha). All the sequences we retrieved belonged to Demospongiae. Similarly, we found no amylase sequence at all in the phylum Ctenophora (*Mnemiopsis leidyi*, *Pleurobrachia pileus*), the phylogenetic position of which is controversial: it has been recovered as the most basal metazoan (WHELAN *et al.* 2017), as Cnidaria's sister group e.g. (PHILIPPE *et al.* 2009; SIMION *et al.* 2017), re-establishing Coelenterata, and also as the earliest branch in the Eumetazoa (animals with a digestive cavity and/or extra cellular digestion) e.g. (PISANI *et al.* 2015).

GH13_1 sequences retrieved from bilaterian animals

The surprising finding of this study, on which we will focus our attention, is the consistent, albeit sparse, occurrence of GH13_1 alpha-amylase sequences in several bilaterian phyla: hemichordates, which are deuterostomes, brachiopods and phoronids (Brachiozoa) and some molluscs and annelids (Eutrochozoa), which are all protostomes. In the brachiopod *Lingula anatina*, two paralogs were found, as in the phoronid *Phoronis australis* (Table 1). In both species, the two copies are located on different contigs. The paralog sequences are rather divergent, given their positions in the tree (Fig 1) and each paralog groups the two species together. This indicates that not only duplication, but also the divergence between paralogs is ancestral to these species, dating back at least to basal Cambrian, according to the TimeTree database (KUMAR *et al.* 2017). GH13_1 sequences were found in other brachiopods as sequence reads (SRA) from transcriptome data only, with no available genomic support (listed in Table 1). We must be cautious with transcriptome data alone, since BLAST hits pertaining to transcripts from contaminating symbionts or parasites may be found (BORNER AND BURMESTER 2017). However, six different brachiopod species returned positive hits, giving some robustness to our finding. Importantly, the related phyla Bryozoa and Nemertea (KOCOT 2015; LUO *et al.* 2018), but see (MARLÉTAZ *et al.* 2019) returned no GH13_1 hits, but these animals are still poorly

represented in sequence databases (only one whole genome in Genbank for Nemertea as of April 2019).

Similarly, we found three gene copies in the genomes of the hemichordates *Sacoglossus kowalevskii* and *Ptychodera flava*. In both species, two copies are close to each other (XP_006816581 and XP_006816582 in *S. kowalevskii*, and their counterparts in *P. flava*) as shown by the topology of the gene tree (Fig. 1). This could suggest independent gene duplication in each species. However, we observed that the two duplicates were arranged in tandem in both species, which would rather suggest concerted evolution of two shared copies. In *P. flava*, this genome region is erroneously annotated as a single gene at the OIST Marine Genomics database. The three copies were therefore probably already present before the split of the two lineages, some 435 mya (KUMAR *et al.* 2017). The third paralog is very divergent from the two other copies, so its divergence from the ancestral copy probably occurred before the species split, as well. Another hemichordate species, *Schizocardium californicum*, harbors a GH13_1 gene, as shown by SRA search in GenBank (Table 1). A positive result was also retrieved from the genome of *Glandiceps talaboti* (Héctor Escrivà, Oceanology Observatory at Banyuls-sur-mer, personal communication). All the species mentioned above belong to the subphylum Enteropneusta. No data from other hemichordate subphyla were available to us.

In molluscs, we found BLAST hits with significant e-values in a few gastropod species from two clades only, the Vetigastropoda (e.g. the abalone *Haliotis* sp.) and the Caenogastropoda (e.g. Ampullariidae such as *Pomacea canaliculata*). We consistently found one copy in eight species belonging to the family Ampullariidae. In *P. canaliculata*, the genome of which has been annotated, the GH13_1 sequence (XP_025109323) lies well inside a 26 Mb long scaffold (linkage group 10, NC_037599) and is surrounded by *bona fide* molluscan genes (Table S1). A GH13_1 sequence was found in the gastropod *Colubraria reticulata*, but in a very short, intronless contig (GenBank accession number CVMW01047267), that was barely longer than the gene itself. Therefore, we disregarded this hit. We also found GH13_1 sequences in a few bivalves, belonging to Mytiloidea (e.g. the mussel *Mytilus galloprovincialis*), Pterioidea (e.g. the pearl oyster *Pinctada imbricata*) and Arcoidea (e.g. *Scapharca broughtoni*). Reciprocal BLAST in GenBank nr using these molluscan high-scoring segment pairs (HSPs) always returned *Lingula anatina* as the best hit. Although several genomes or transcriptomes have been sequenced from other bivalve and gastropod clades, we retrieved no GH13_1-like sequences from the databases except in the aforementioned clades. We found no such sequence in cephalopods either. In some cases, the sequences were retrieved from the TSA database (see

Table 1), whose issues were mentioned above. As an example, we found a significant hit in a transcriptome database of the sea hare *Aplysia californica* (TSA GBDA01069500) but this sequence was not found in the *A. californica* genome, which is well annotated; it was indeed related to ciliates.

In annelids, we found occurrences of GH13_1 genes in a few species, the genomes of which are still not fully assembled, namely the "polychaetes" *Hydroides elegans* and *Pygospio elegans* but not in the well annotated genome of *Capitella teleta.* We also recovered HSPs from the clitellate *Glossoscolex paulistus* but not from *Amynthas corticis* or *Eisenia fetida*. Therefore, in molluscs as well as in annelids, the presence of GH13_1 genes is scattered across lineages. We found that the mollusc GH13_1-like sequences were much shorter, either truncated at the C-terminal, or this region was so divergent from the query sequence (*L. anatina)* that it was impossible to identify, assemble and align it with our data set (Fig. S1). In addition, we found that the annelid *Hydroides elegans* had an internal deletion, which precluded its inclusion in the phylogenetic analysis. This suggests that those sequences may not have alpha-amylase activity.

Gene tree analysis: position of bilaterian sequences

The goal of the gene tree analysis is to examine whether the occurrence of GH13_1 genes in bilaterian animals may be due to horizontal transfer (HGT) or if they descend from a GH13_1 alpha-amylase copy ancestral to Unikonts. In the first case, the bilaterians GH13_1 sequences are unlikely to cluster together and the gene tree topology will likely display one or more nodes that are inconsistent with the bilaterian phylogeny. In the second case, the bilaterian sequences are expected to recover a bilaterian clade and to have a cnidarian clade as its sister group (LAUMER *et al.* 2018). The actual tree topology (Fig. 1) is not that straightforward when it comes to the bilaterian relationships, although we may rule out any proximity of bilaterians GH13_1 sequences with unicellular or fungal sequences, regardless of tree rooting.

All Cnidarian orthologs form a well-supported cluster. The sistership between Corallimorpharia and Scleractinia reflects what was recovered in species trees using different markers (e.g. (RODRIGUEZ *et al.* 2014)), although the Scleractinia topology disagrees with previous phylogenetic analyses of the order (e.g. (BARBEITOS *et al.* 2010)). The other cluster within Cnidaria is mainly composed of actiniarian (sea anemone) sequences, but it also includes the sequence of *Renilla reniformis,* which belongs to Octocorallia, a clade that is sister to Scleractinia + Corallimorpharia (RODRIGUEZ *et al.* 2014). This strong inconsistency between the

R. reniformis position in the GH13_1 topology and the species tree topology may be interpreted as due to a horizontal transfer event that would have occurred within Cnidaria. Most bilaterian sequences are clustered with Cnidaria, as phylogenetically expected in the case of a shared ancestral gene, as a robust cluster grouping one Brachiozoa (brachiopod/phoronid) copy, the molluscs and the annelids, which is consistent with the phylogeny. However, the tandem hemichordate duplicates and the other Brachiozoa genes are not included in the bilaterian clade, but remain ingroup relative to the sponge sequences.

Interestingly, the two remaining hemichordate sequences are the earliest diverging lineage of the Metazoa + Choanoflagellata cluster, since they are branched with the placozoan *Trichoplax adhaerens* sequence, this relationship being strongly supported whatever the tree reconstruction method employed (not shown - Fig. 1). In order to check for the possibility of a long branch attraction (LBA), which would artificially cluster hemichordate and placozoan sequences, we performed Tajima's relative rate tests (TAJIMA 1993) using MEGA7 (KUMAR *et al.* 2016). The sequence of *S. kowalevskii* XP_006819810, suspected to evolve fast, was compared with its paralog XP_006816581, using five different outgroups, i.e. the three sponges and the two choanoflagellates. Unexpectedly, the χ^2 tests returned non-significant values in two tests and significant values in three tests (Table S2). Therefore, with our data, LBA cannot be entirely ruled out in this particular case.

Analysis of intron positions

Intron positions may be valuable markers when reconstituting gene histories. We identified 56 intron positions from the subset of species of the general tree for which we could find data (Fig. 2). Only one intron position is widely shared among these GH13_1 gene sequences. It is the first position reported in the alignment, and it lies just upstream to the first conserved part of the alignment. The main observation is the numerous conserved positions across bilaterian sequences (10 positions), and between bilaterian sequences and the sponge and the Placozoa (7 positions). In addition, three positions are common to bilaterians and the choanoflagellate *Monosiga brevicollis*. In contrast, the Cnidaria have few introns, with positions different from the sponge and the bilaterians, except for position 1. The other species under examination, i.e. protists and fungi, have essentially specific intron positions. This is a further argument to state that the occurrence of GH13_1 alpha-amylases in some bilaterians is a story that is internal to metazoans.

Discussion

We have shown here that a limited number of bilaterian animals, all being aquatic species, namely hemichordates, brachiopods and phoronids, and some sparse molluscs and annelids, do have GH13_1 alpha-amylase genes. Note that all those species do have at least one "classical" animal alpha-amylase of the GH13_15/24 subfamilies. We are quite confident that the GH13_1 sequences we found are not due to contaminating DNA. First, two species with whole genome sequenced and assembled were found to harbor such genes in each phylum Hemichordata and Brachiozoa, and the mollusc *Pomacea canaliculata* also has a well annotated genome. Additional sequences from other species belonging to these phyla were gathered from sketchy data, i.e. low-quality assembled genomes, transcriptomes or sequence read archive databases, which added some support to the presence of these amylase genes. Although transcriptome and rough genomic data should be handled with care, this lends support to our observations. Moreover, reciprocal BLAST from the transcriptome hits always returned a bilaterian (*L. anatina* or *S. kowalevskii*) best hit, not fungal, protist or other non-bilaterian GH13_1 sequence. Second, the bilaterian sequences retrieved from assembled genomes were inside long contigs, and mostly surrounded by genes showing bilaterian best BLAST hits (Table S1). However, the *S. kowalevskii* XP_006819810 gene could appear somewhat dubious, since it is placed at the distal end of a contig, with only two other genes on the contig (Table S1), one of which has a placozoan best hit. But its *P. flava* counterpart is well inside a very gene-rich contig. Therefore, these seemingly non-bilaterian genes are well in bilaterian genomic contexts.

The evolutionary scenario proposed by Da Lage et al (Da Lage *et al.* 2007) suggested that the GH13_1 alpha-amylase gene ancestral to Unikonts (Amoebozoa and Opisthokonts, i.e. Fungi and Metazoa/Choanoflagellata) was totally absent from Bilateria, due to complete replacement by a new alpha-amylase, originating from a bacterium through horizontal gene transfer (HGT). The new data unveils a more complicated story. There are two explanations which are mutually exclusive. The first explanation is that several HGTs occurred from non-bilaterian to hemichordate and Lophotrochozoa ancestors. The second explanation is that the ancestral gene was not lost in all bilaterian lineages, but remained (given the current data) in hemichordates, brachiopods and phoronids, and in scattered lineages across Mollusca and Annelida.

The hypothesis of HGT requires several such events between metazoans. It implies that HGTs obviously happened after the split of the two main branches of bilaterians, protostomes and deuterostomes, otherwise the transferred copies should have been lost in most phyla, like in the alternative hypothesis. More precisely, in the case of Lophotrochozoa, this would have occurred before the diversification of this clade and after its divergence from the Platyzoa, some 700 mya(KUMAR *et al.* 2017); in the case of hemichordates, after diverging from their common ancestor with the echinoderms, and before the divergence between *S. kowalevskii* and *Ptychodera flava*, i.e. between 657 and ca. 435 mya (KUMAR *et al.* 2017). Therefore, we may infer *at least* two HGTs, early in the evolution of each phylum, with subsequent losses in Lophotrochozoa (Fig. 3). The donor species, given the sequence clustering in the trees, could be related to cnidarians. However, we have underlined that the intron-exon structures of the bilaterian sequences were most similar to the one of the sponge, and that the cnidarian GH13_1 amylases had very different structures. This may be possible if the donors were related to cnidarians, perhaps an extinct phylum or an ancestor of extant Cnidaria, but had conserved the ancestral structures exemplified by the sponge and the placozoan. Indeed, if the structure shared by the sponge, the placozoan and the bilaterians reflects the ancestral state, cnidarians must have undergone a drastic rearrangement of the intron-exon structure of this gene. This would be in line with the long internal branch leading to this clade in the trees (Fig. 1), which suggests accelerated evolution.

The alternative hypothesis of massive GH13_1 gene loss in most phyla except the ones where we found such sequences seems no more parsimonious. It requires many losses, depending on the phylogeny used (Fig. 3). For instance, regarding deuterostomes, one loss occurred in echinoderms and the other in the chordates. In protostomes, GH13_1 loss in ecdysozoans, and independently in several lophotrochozoan lineages would be required to produce the observed pattern.

Although not parsimonious in terms of number of events, we would favor the gene loss hypothesis, because this is a common phenomenon, especially given how ubiquitous co-option is (HEJNOL AND MARTINDALE 2008; FLORES AND LIVINGSTONE 2017). In this respect, the GH13_15/24 gene that was acquired from a bacterium is a type of horizontal transfer akin to what Husnik and McCutcheon called a “maintenance transfer” since it allowed the original function to be maintained while the primitive gene became free to evolve or even to be lost (HUSNIK AND MCCUTCHEON 2018) (see also (DA LAGE *et al.* 2013)). In contrast, while numerous cases of HGT from bacteria to metazoans have been reported (e.g., (DUNNING HOTOPP 2011;

HAEGEMAN *et al.* 2011; WYBOUV *et al.* 2016; CORDAUX AND GILBERT 2017)), very few HGT events have been inferred that involve a metazoan donor and a metazoan receiver (GRAHAM *et al.* 2012; GASMI *et al.* 2015). Thus, our current knowledge on HGT suggests that this type of transfer might be very rare between metazoans, and that two or more such events would be quite unlikely to explain the current taxonomic distribution of metazoan GH13_1 genes. In addition, it has been shown that a seemingly patchy gene distribution suggestive of HGT may, after more comprehensive taxon sampling, turn out to be rather due to recurrent gene losses (HUSNIK AND MCCUTCHEON 2018). The conservation across phyla of the intron-exon structure, probably ancestral to the metazoans, would not be surprising (SULLIVAN *et al.* 2006 ; SRIVASTAVA *et al.* 2008; SRIVASTAVA *et al.* 2010). For instance, 82% of human introns have orthologous introns in *T. adhaerens* (SRIVASTAVA *et al.* 2008).

In this work, the different sequences were assumed to be alpha-amylases according to BLAST e-values only. In addition, we also assumed that they all belong to the GH13_1 subfamily. Indeed, some of them have been assigned to this subfamily in the reference database CAZy.org (see Table 1), and if we add sequences from the closest subfamilies, namely GH13_2 or GH13_19 (STAM *et al.* 2006) in the alignment and in the phylogenetic tree, the putative GH13_1 and the ascertained GH13_1 remain well clustered together (not shown). It is possible that modifications of a few amino acid positions could bring a change in the substrate or catalytic activity. For instance, concerning the substrate affinity, when the genome of *L. anatina* was released, the authors hypothesized a biomineralization pathway that involves acid proteins, as found in scleractinians and molluscs (MARIN *et al.* 2007; RAMOS-SILVA *et al.* 2013). Given the calcium binding activity of alpha-amylases (BOEL *et al.* 1990; GROSSMAN AND JAMES 1993; SVENSSON 1994; PUJADAS AND PALAU 2001), the presence of both GH13_1 and GH13_15 subfamilies in *L. anatina* opens the possibility for the neofunctionalization of one of them in the biomineralization process. In the analyses performed by those authors, no amylase was found in the shell matrix, but this does not exclude the possibility of its presence in the pathway. The fact that in molluscs, the sequences are incomplete compared to the brachiopod query or to the sponge and cnidarian GH13_1 amylases, and therefore probably devoid of an amylolytic function, would add credence to another function. This conjecture requires further investigation. On the other hand, the full-size GH13_1 sequences only present in a few bilaterians could have remained true alpha-amylases with the classical function, but this would make even more enigmatic why they have been conserved, either by descent or by horizontal transfer.

Acknowledgments: We want to thank Pedro E. Vieira for introducing AD to molecular analyses. We are grateful to Didier Casane and Emmanuelle Renard for fruitful advise and discussion. We thank Héctor Escrivà for sharing sequence data. This work was funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (process no. 141565/2017-9) to AD and regular funding of the CNRS to JLDL and CG. CG is also supported by a grant from Agence Nationale de la Recherche (ANR-15-CE32-0011-01 TransVir).

References :

Anisimova, M., and O. Gascuel, 2006 Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic Biology 55**:** 539-552.

Barbeitos, M. S., S. L. Romano and H. R. Lasker, 2010 Repeated loss of coloniality and symbiosis in scleractinian corals. Proc Natl Acad Sci U S A 107**:** 11877-11882.

Boel, E., L. Brady, A. M. Brzozoxski, Z. Derewenda, G. G. Dodson *et al.*, 1990 Calcium-binding in α-amylases: an X-ray diffraction study at 2.1 Â resolution of two enzymes from *Aspergillus*. Biochemistry 29**:** 6244-6249.

Borner, J., and T. Burmester, 2017 Parasite infection of public databases: a data mining approach to identify apicomplexan contaminations in animal genome and transcriptome assemblies. BMC Genomics 18**:** 100.

Cordaux, R., and C. Gilbert, 2017 Evolutionary significance of Wolbachia-to-animal horizontal gene transfer: Female sex determination and the f element in the isopod *Armadillidium vulgare*. Genes 8.

Da Lage, J.-L., M. Binder, A. Hua-Van, S. Janeček and D. Casane, 2013 Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial alpha-amylase gene to Basidiomycetes. BMC Evolutionary Biology 13**:** 40.

Da Lage, J.-L., E. G. J. Danchin and D. Casane, 2007 Where do animal α-amylases come from? An interkingdom trip. FEBS Letters 581**:** 3927-3935.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet *et al.*, 2008 Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36(Web Server Issue)**:** W465-469.

Dunning Hotopp, J. C., 2011 Horizontal gene transfer between bacteria and animals. Trends in Genetics 27.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32**:** 1792-1797.

Flores, R. L., and B. T. Livingstone, 2017 The skeletal proteome of the sea star *Patiria miniata* and evolution of biomineralization in echinoderms. BMC Evolutionary Biology 17**:** 1-14.

Gasmi, L., H. Boulain, A. Hua-Van, K. Musset, A. K. Jakubowska *et al.*, 2015 Recurrent domestication by Lepidoptera of genes from their parasites mediated by Bracoviruses. PLoS Genetics 11**:** e1005470.

Graham, L. A., J. Li, W. S. Davidson and P. L. Davies, 2012 Smelt was the likely beneficiary of an antifreeze gene laterally transferred between fishes. BMC Evolutionary Biology 12.

Grossman, G. L., and A. A. James, 1993 The salivary glands of the vector mosquito Aedes aegypti, express a novel member of the amylase gene family. Insect Molecular Biology 1**:** 223-232.

Guindon, S., and O. Gascuel, 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. . Systematic Biology 52**:** 696-704.

Haegeman, A., J. T. Jones and E. G. Danchin, 2011 Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Molecular Plant and Microbe Interactions 24**:** 879-887.

Hejnol, A., and M. Q. Martindale, 2008 Acoel development indicates the independent evolution of the bilaterian mouth and anus. Nature 456**:** 382-386.

Henrissat, B., and G. Davies, 1997 Structural and sequence-based classification of glycoside hydrolases. Current Opinion in Structural Biology 7**:** 637-644.

Husnik, P., and J. P. McCutcheon, 2018 Functional horizontal gene transfer from bacteria to eukaryotes. Nature reviews Microbiology 16**:** 67-79.

Ip, J. C. H., H. Mu, Q. Chen, J. Sun, S. Ituarte *et al.*, 2018 AmpuBase: a transcritome database for eight species of apple snails (Gastropoda: Ampullariidae). BMC Genomics 19**:** 179.

Kocot, K. M., 2015 On 20 years of Lophotrochozoa. Organisms Diversity and Evolution 16**:** 329-343.

Kocot, K. M., T. H. Struck, J. Merkel, D. S. Waits, C. Todt *et al.*, 2017 Phylogenomics of Lophotrochozoa with considerations of systematic error. Systematic Biology 66**:** 256-282.

Kumar, S., G. Stecher, M. Suleski and S. B. Hedges, 2017 TimeTree: a resource for timelines, timetrees, and divergence times. Molecular Biology and Evolution 34**:** 1812-1819.

Kumar, S., G. Stecher and K. Tamura, 2016 MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. Molecular Biology and Evolution 33**:** 1870-1874.

Laumer, C. E., H. Gruber-Vodicka, M. G. Hadfield, V. B. Pearse, A. Riesgo *et al.*, 2018 Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. eLife 7**:** e36278.

Letunic, I., and P. Bork, 2016 Interactive tree of life(iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research 44**:** W242-245.

Lombard, V., H. HGolaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, 2014 The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research 42**:** 490-495.

Luo, Y.-J., M. Kanda, R. Koyanagi, K. Hisata, T. Akiyama *et al.*, 2018 Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. Nature Ecology and Evolution 2**:** 141-151.

Luo, Y.-J., T. Takeuchi, R. Koyanagi, L. Yamada, M. Kanda *et al.*, 2015 The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. Nature communications 6**:** 8301.

MacGregor, E. A., S. Janecek and B. Svenssson, 2001 Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochimica et Biophysica Acta 1546**:** 1-20.

Marin, F., G. Luquet, B. Marie and D. Medakovic, 2007 Molluscan shell proteins: primary structure, origin, and evolution. Current Topics in Developmental Biology 80**:** 209-276.

Marlétaz, F., K. T. C. A. Peljnenburg, T. Goto, N. Satoh and D. S. Rokhsar, 2019 A New spiralian phylogeny places the enigmatic arrow worms among Gnathiferans. Current Biology 29**:** 312-318.

Philippe, H., R. Derelle, P. Lopez, K. Pick, C. Borchiellini *et al.*, 2009 Phylogenomics revives traditional views on deep animal relationships. Current Biology 19**:** 706-712.

Pisani, D., W. Pett, M. Dohrmann, R. Feuda, O. Rota-Stabelli *et al.*, 2015 Genomic data do not support comb jellies as the sister group to all other animals. Proc Natl Acad Sci U S A 112**:** 15402-15407.

455 Plazzi, F., A. Ceregato, M. Taviani and M. Passamonti, 2011 A molecular phylogeny of bivalve
456 mollusks: Ancient radiations and divergences as revealed by mitochondrial genes. PLoS
457 One.
458 Pujadas, G., and J. Palau, 2001 Evolution of α-amylase: architectural features and key residues
459 in the stabilization of the (b/a)8 scaffold. Molecular Biology and Evolution 18**:** 38-54.
460 Ramos-Silva, P., J. Kaandorp, L. Huisman, B. Marie, I. Zanella-Cléon *et al.*, 2013 The skeletal
461 proteome of the coral *Acropora millepora*: the evolution of calcification by co-option
462 and domain shuffling. Molecular Biology and Evolution 30**:** 2099-2112.
463 Rodríguez, E., M. S. Barbeitos, M. R. Brugler, L. M. Crowley, A. Grajales *et al.*, 2014 Hidden
464 among Sea Anemones: The First Comprehensive Phylogenetic Reconstruction of the
465 Order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) Reveals a Novel Group of
466 Hexacorals. PLoS One 9**:** e96998.
467 Simion, P., H. Philippe, D. Baurain, M. Jager, D. J. Richter *et al.*, 2017 A large and consistent
468 phylogeny dataset supports sponges as the sister group to all other animals. Current
469 Biology 27**:** 958-967.
470 Srivastava, M., E. Begovic, J. Chapman, N. H. Putnam, U. Hellsten *et al.*, 2008 The *Trichoplax*
471 genome and the nature of placozoans. Nature 454**:** 955-960.
472 Srivastava, M., O. Simakov, J. Chapman, B. Fahey, M. E. A. Gauthier *et al.*, 2010 The
473 Amphimedon queenslandica genome and the evolution of animal complexity. Nature
474 466**:** 720-726.
475 Stam, M. R., E. G. J. Danchin, C. Rancurel, P. M. Coutinho and B. Henrissat, 2006 Dividing
476 the large glycoside hydrolase family 13 into subfamilies: towards improved functional
477 annotations of α-amylase-related proteins. Protein Engineering, Design & Selection 19**:**
478 555-562.
479 Sullivan, J. C., A. M. Reitzel and J. R. Finnerty, 2006 A high percentage of introns in human
480 genes were present early in animal evolution: evidence from the basal metazoan
481 *Nematostella vectensis*. Genome Informatics 17**:** 219-229.
482 Svensson, B., 1994 Protein engineering in the α-amylase family: catalytic mechanism, substrate
483 specificity, and stability. Plant Molecular Biology 25**:** 141-157.
484 Tajima, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis.
485 Genetics 135**:** 599-607.
486 Uribe, J., Y. Kano, J. Templado and R. Zardoya, 2016 Mitogenomics of Vetigastropoda:
487 insights into the evolution of pallial symmetry. Zoologica Scripta 45**:** 145-159.

Whelan, N. V., K. M. Kocot, T. P. Moroz, K. Mukherjee, P. Williams *et al.*, 2017 Ctenophore relationships and their placement as the sister group to all other animals. Nature Ecology and Evolution 1**:** 1737-1746.

Whelan, S., and N. Goldman, 2001 A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution 18**:** 691-699.

Wybouv, N., Y. Pauchet, D. G. Heckel and T. Van Leuuwen, 2016 Horizontal gene transfer contributes to the evolution of arthropod herbivory. Genome Biology and Evolution 8**:** 3594-3613.

Legends of figures :

Figure 1: ML tree of GH13_1 protein sequences of metazoan and non-metazoan species. The tree was rooted by placing fungi and unicellular organisms, except choanoflagellates, as outgroups. The numbers at the nodes are the aLRT supports. Dark green: hemichordates; light blue: brachiozoans; red: cnidarians, dark blue: sponges; orange: placozoans; pink: choanoflagellates; purple: amoebozoans; brown: fungi; grey, molluscs; bright green: annelids; black: other protists.

Figure 2: Intron positions compared across the sampled GH13_1 genes. The intron positions found in the studied parts of the sequences were numbered from 1 to 56. Pink: phase zero introns; green: phase 1 introns; blue: phase 2 introns. The black horizontal bar separates bilaterians from species where GH13_1 alpha-amylases are considered native. The color code for species is the same as in Figure 1.

Figure 3: Two scenarii of HGT/gene losses of the GH13_1 genes. HGT or gene loss events were plotted on one of the proposed phylogenies of Bilateria, adapted from references (PLAZZI *et al.* 2011; KOCOT 2015; LUO *et al.* 2015; URIBE *et al.* 2016; KOCOT *et al.* 2017; LUO *et al.* 2018). A: HGT hypothesis. Black lozanges represent the HGT events, crosses indicate subsequent GH13_1 loss events. B: Gene loss hypothesis. Crosses indicate GH13_1 loss events. Taxa for which all available genomes were found to contain one or more GH13_1 sequences are in red; taxa in

519 which GH13_1 sequences were not found are in black; taxa for which only a fraction of
520 available genomes were found to contain a GH13_1 sequence are in orange. Divergence times
521 are from (KUMAR *et al.* 2017).

522

523

524 **Legends of Supplementary figures**

525

526

527 **Figure S1 :** General protein alignment of the GH13_1 sequences used for the gene tree.
528 Background colors indicate the conservation level of amino acids : red : 100% identity; orange :
529 80 to 100% identity ; yellow : 60 to 80% identity ; white : below 60% identity.

530

531

532

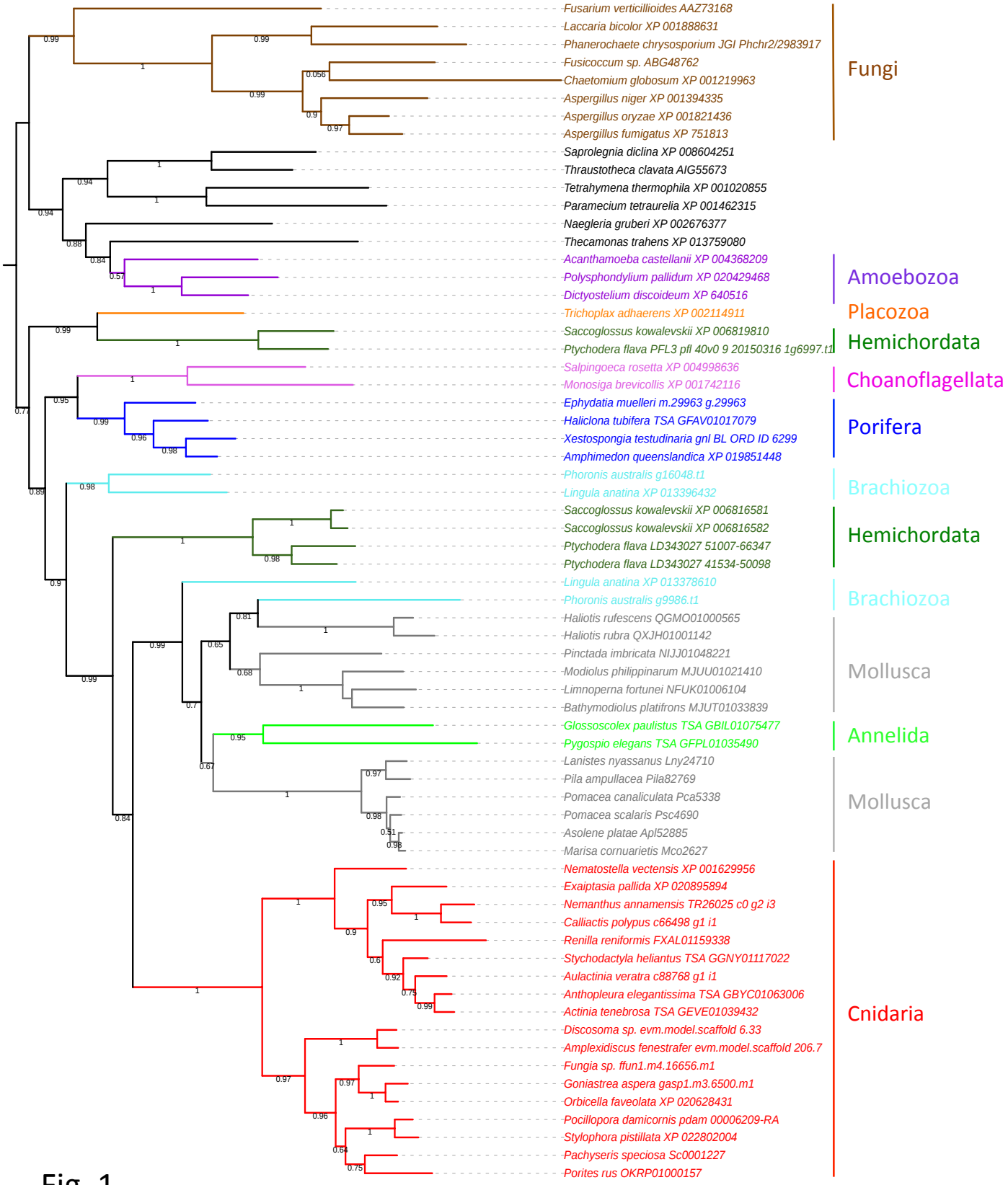

Fusarium verticillioides AAZ73168
Laccaria bicolor XP 001888631
Phanerochaete chrysosporium JGI Phchr2/2983917
Fusicoccum sp. ABG48762
Chaetomium globosum XP 001219963
Aspergillus niger XP 001394335
Aspergillus oryzae XP 001821436
Aspergillus fumigatus XP 751813
Fungi
Saprolegnia diclina XP 008604251
Thraustotheca clavata AIG55673
Tetrahymena thermophila XP 001020855
Paramecium tetraurelia XP 001462315
Naegleria gruberi XP 002676377
Thecamonas trahens XP 013759080
Acanthamoeba castellanii XP 004368209
Polysphondylium pallidum XP 020429468
Dictyostelium discoideum XP 640516
Amoebozoa
Trichoplax adhaerens XP 002114911
Placozoa
Saccoglossus kowalevskii XP 006819810
Ptychodera flava PFL3 pfl 40v0 9 20150316 1g6997.t1
Hemichordata
Salpingoeca rosetta XP 004998636
Monosiga brevicollis XP 001742116
Choanoflagellata
Ephydatia muelleri m.29963 g.29963
Haliclona tubifera TSA GFAV01017079
Xestospongia testudinaria gnl BL ORD ID 6299
Amphimedon queenslandica XP 019851448
Porifera
Phoronis australis g16048.t1
Lingula anatina XP 013396432
Brachiozoa
Saccoglossus kowalevskii XP 006816581
Saccoglossus kowalevskii XP 006816582
Ptychodera flava LD343027 51007-66347
Ptychodera flava LD343027 41534-50098
Hemichordata
Lingula anatina XP 013378610
Phoronis australis g9986.t1
Brachiozoa
Haliotis rufescens QGMO01000565
Haliotis rubra QXJH01001142
Pinctada imbricata NIJJ01048221
Modiolus philippinarum MJUU01021410
Limnoperna fortunei NFUK01006104
Bathymodiolus platifrons MJUT01033839
Mollusca
Glossoscolex paulistus TSA GBIL01075477
Pygospio elegans TSA GFPL01035490
Annelida
Lanistes nyassanus Lny24710
Pila ampullacea Pila82769
Pomacea canaliculata Pca5338
Pomacea scalaris Psc4690
Asolene platae Apl52885
Marisa cornuarietis Mco2627
Mollusca
Nematostella vectensis XP 001629956
Exaiptasia pallida XP 020895894
Nemanthus annamensis TR26025 c0 g2 i3
Calliactis polypus c66498 g1 i1
Renilla reniformis FXAL01159338
Stychodactyla heliantus TSA GGNY01117022
Aulactinia veratra c88768 g1 i1
Anthopleura elegantissima TSA GBYC01063006
Actinia tenebrosa TSA GEVE01039432
Discosoma sp. evm.model.scaffold 6.33
Amplexidiscus fenestrafer evm.model.scaffold 206.7
Fungia sp. ffun1.m4.16656.m1
Goniastrea aspera gasp1.m3.6500.m1
Orbicella faveolata XP 020628431
Pocillopora damicornis pdam 00006209-RA
Stylophora pistillata XP 022802004
Pachyseris speciosa Sc0001227
Porites rus OKRP01000157
Cnidaria

Fig. 1

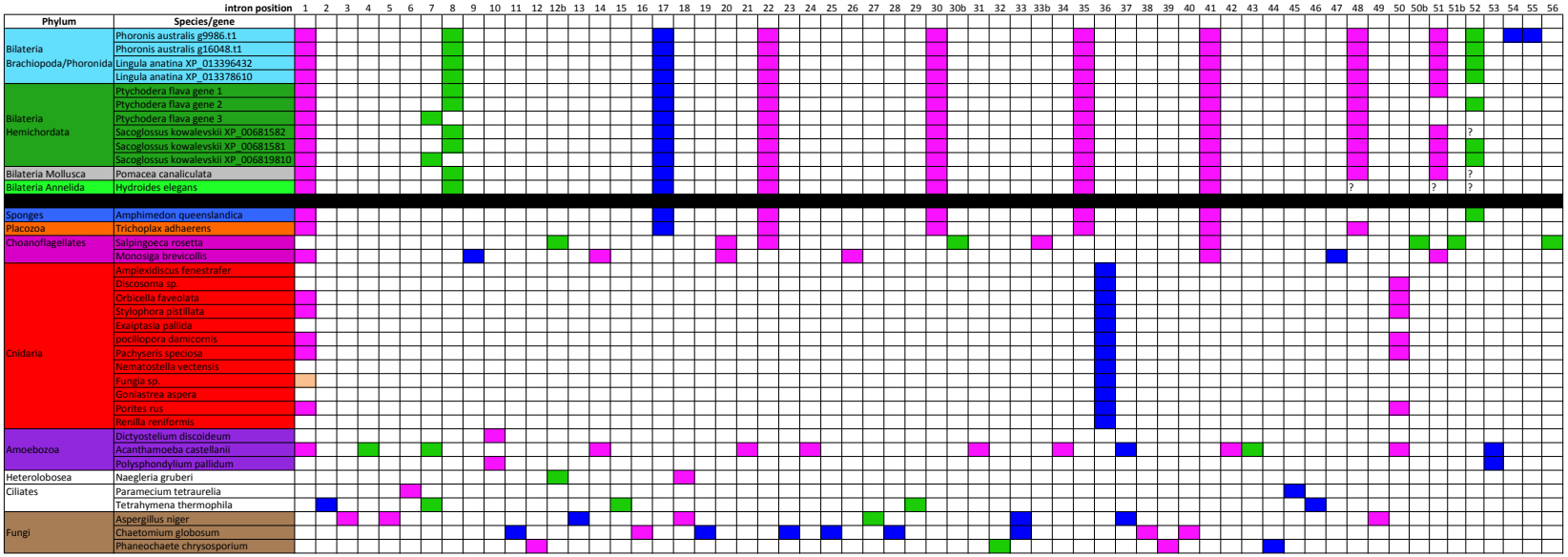
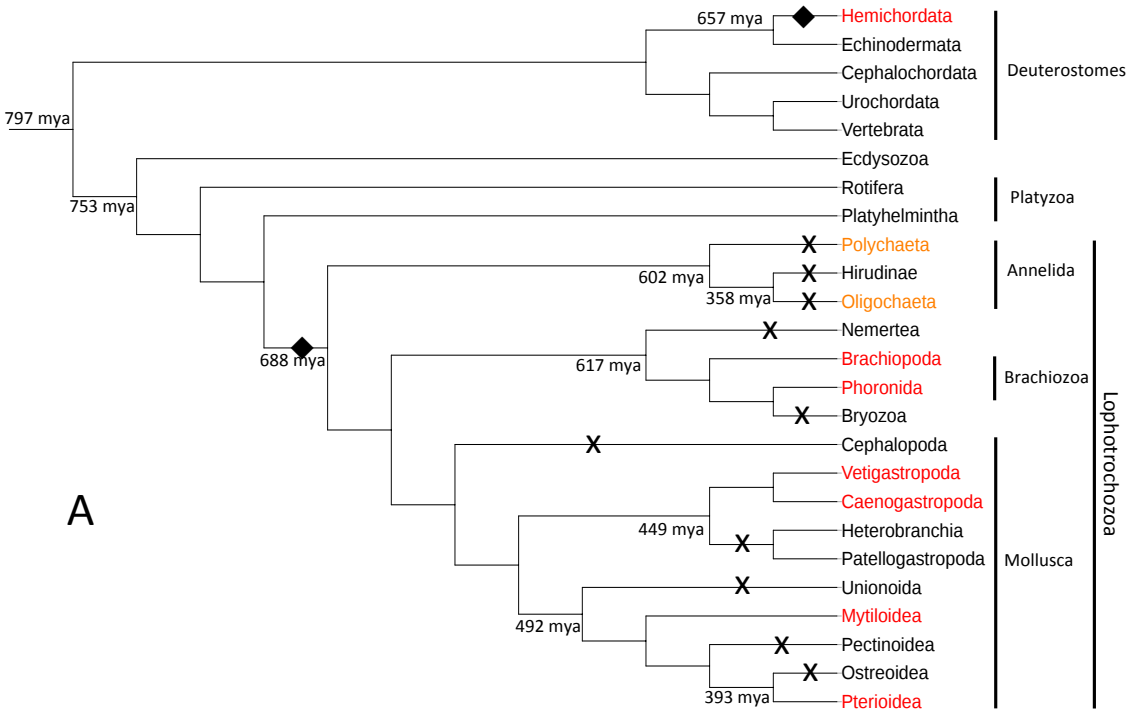

intron position 1 2 3 4 5 6 7 8 9 10 11 12 12b 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 30b 31 32 33 33b 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 50b 51 51b 52 53 54 55 56
Phylum
Species/gene
Bilateria
Brachiopoda/Phoronida
Phoronis australis g9986.t1
Phoronis australis g16048.t1
Lingula anatina XP_013396432
Lingula anatina XP_013378610
Bilateria
Hemichordata
Ptychodera flava gene 1
Ptychodera flava gene 2
Ptychodera flava gene 3
Sacoglossus kowalevskii XP_00681582
Sacoglossus kowalevskii XP_00681581
Sacoglossus kowalevskii XP_006819810
Bilateria Mollusca
Pomacea canaliculata
Bilateria Annelida
Hydroides elegans
Sponges
Amphimedon queenslandica
Placozoa
Trichoplax adhaerens
Choanoflagellates
Salpingoeca rosetta
Monosiga brevicollis
Cnidaria
Amplexidiscus fenestrafer
Discosoma sp.
Orbicella faveolata
Stylophora pistillata
Exaiptasia pallida
pocillopora damicornis
Pachyseris speciosa
Nematostella vectensis
Fungia sp.
Goniastrea aspera
Porites rus
Renilla reniformis
Amoebozoa
Dictyostelium discoideum
Acanthamoeba castellanii
Polysphondylium pallidum
Heterolobosea
Naegleria gruberi
Ciliates
Paramecium tetraurelia
Tetrahymena thermophila
Fungi
Aspergillus niger
Chaetomium globosum
Phaneochaete chrysosporium

Fig. 2


657 mya
Hemichordata
Echinodermata
Cephalochordata
Urochordata
Vertebrata
Deuterostomes
797 mya
Ecdysozoa
Rotifera
Platyhelmintha
Platyzoa
753 mya
Polychaeta
Hirudinae
602 mya
358 mya
Oligochaeta
Annelida
Nemertea
688 mya
617 mya
Brachiopoda
Phoronida
Bryozoa
Brachiozoa
Lophotrochozoa
Cephalopoda
Vetigastropoda
Caenogastropoda
449 mya
Heterobranchia
Patellogastropoda
Mollusca
Unionoida
Mytiloidea
492 mya
Pectinoidea
Ostreoidea
393 mya
Pterioidea
A

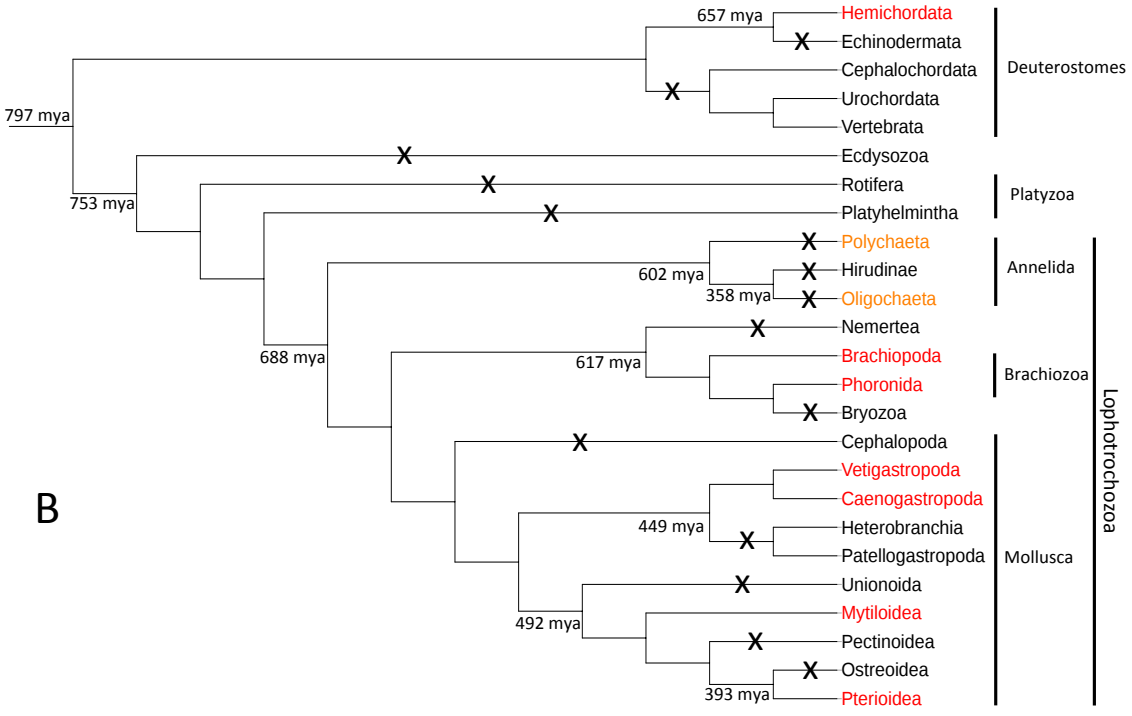

657 mya
Hemichordata
Echinodermata
Cephalochordata
Urochordata
Vertebrata
Deuterostomes
797 mya
Ecdysozoa
Rotifera
Platyhelmintha
Platyzoa
753 mya
Polychaeta
Hirudinae
602 mya
358 mya
Oligochaeta
Annelida
Nemertea
688 mya
617 mya
Brachiopoda
Phoronida
Bryozoa
Brachiozoa
Lophotrochozoa
Cephalopoda
Vetigastropoda
Caenogastropoda
449 mya
Heterobranchia
Patellogastropoda
Mollusca
Unionoida
Mytiloidea
492 mya
Pectinoidea
Ostreoidea
393 mya
Pterioidea
B

Fig. 3

Table 1: GH13_1-like sequences found after BLAST searches in online databases (not comprehensive for unicellars, without the Fungi). * : sequences which have not been characterized as protein-coding, in sequenced genomes with long contigs ; (1) : from short DNA sequences (except Sequence reads archive) ; ** : reported as GH13_1 in CAZy.

Phylum	**Species**	**Database**	**Accession**
NON BILATERIAN METAZOA			
Porifera Demospongiae Heteroscleromorpha	*Amphimedon queenslandica*	GenBank proteins	XP_019851448
Porifera Demospongiae Heteroscleromorpha	*Ephydatia muelleri (1)*	Compagen.org	m.29963 g.29963
Porifera Demospongiae Heteroscleromorpha	*Haliclona tubifera*	GenBank TSA	GFAV01017079
Porifera Demospongiae Heteroscleromorpha	*Spongilla lacustris*	GenBank SRA	SRX470277
Porifera Demospongiae Heteroscleromorpha	*Xestospongia testudinaria (1)*	Reefgenomics.org	gnl|BL_ORD_ID|6299
Cnidaria Hexacorallia Actiniaria	*Actinia tenebrosa*	GenBank TSA	GEVE01039432
Cnidaria Hexacorallia Actiniaria	*Anthopleura elegantissima*	GenBank TSA	GBYC01063006
Cnidaria Hexacorallia Actiniaria	*Anthopleura buddemeieri (1)*	Reefgenomics.org	c117986_g2_i1
Cnidaria Hexacorallia Actiniaria	*Aulactinia veratra (1)*	Reefgenomics.org	c88768_g1_i1
Cnidaria Hexacorallia Actiniaria	*Calliactis polypus (1)*	Reefgenomics.org	c66498_g1_i1
Cnidaria Hexacorallia Actiniaria	*Exaiptasia pallida*	GenBank proteins	XP_020895894
Cnidaria Hexacorallia Actiniaria	*Nematostella vectensis*	GenBank proteins	XP_001629956
Cnidaria Hexacorallia Actiniaria	*Stychodactyla heliantus*	GenBank TSA	GGNY01117022
Cnidaria Hexacorallia Corallimorpharia	*Amplexidiscus fenestrafer**	Reefgenomics.org	evm.model.scaffold_206.7
Cnidaria Hexacorallia Corallimorpharia	*Discosoma sp.**	Reefgenomics.org	evm.model.scaffold_6.33
Cnidaria Hexacorallia Scleratinia	*Acropora digitifera*	GenBank proteins	XP_015760547 partial
Cnidaria Hexacorallia Scleratinia	*Acropora tenuis**	Reefgenomics.org	aten_0.1.m1.10359.m1
Cnidaria Hexacorallia Scleratinia	*Fungia sp.**	Reefgenomics.org	ffun1.m4.16656.m1
Cnidaria Hexacorallia Scleratinia	*Goniastrea aspera**	Reefgenomics.org	gasp1.m3.6500.m1
Cnidaria Hexacorallia Scleratinia	*Nemanthus annamensis (1)*	Reefgenomics.org	TR26025|c0_g2_i3
Cnidaria Hexacorallia Scleratinia	*Orbicella faveolata*	GenBank proteins	XP_020628431
Cnidaria Hexacorallia Scleratinia	*Pachyseris speciosa**	Reefgenomics.org	Sc0001227 74283-80000
Cnidaria Hexacorallia Scleratinia	*Pocillopora damicornis**	Reefgenomics.org	pdam_00006209-RA
Cnidaria Hexacorallia Scleratinia	*Porites lutea**	Reefgenomics.org	plut2.m8.18618.m1
Cnidaria Hexacorallia Scleratinia	*Porites rus*	GenBank genomes	OKRP01000157

Cnidaria Hexacorallia Scleratinia	*Stylophora pistillata*	GenBank proteins	XP_022802004
Cnidaria Octocorallia Pennatulacea	*Renilla reniformis**	GenBank genomes	FXAL01159338
Placozoa	*Trichoplax adhaerens*	GenBank proteins	XP_002114911
BILATERIA			
Brachiopoda Linguliformea	*Glottidia pyramidata*	GenBank SRA	SRX731468 (transcriptome)
Brachiopoda Linguliformea	*Lingula anatina*	GenBank proteins	XP_013396432
Brachiopoda Linguliformea	*Lingula anatina*	GenBank proteins	XP_013378610
Brachiopoda Phoroniformea or Phoronida	*Phoronis australis*	marinegenomics	g9986.t1
Brachiopoda Phoroniformea or Phoronida	*Phoronis australis*	marinegenomics	g16048.t1
Brachiopoda Phoroniformea or Phoronida	*Phoronopsis harmeri*	GenBank SRA	SRX1121914 (transcriptome)
Brachiopoda Craniiformea	*Novocrania anomala*	GenBank SRA	SRX731472 (transcriptome)
Brachiopoda Rhynchonelliformea	*Kraussina rubra*	GenBank SRA	SRX112037 (transcriptome)
Brachiopoda Rhynchonelliformea	*Macandrevia cranium*	GenBank SRA	SRX731471 (transcriptome)
Brachiopoda Rhynchonelliformea	*Terebratalia transversa*	GenBank SRA	SRX1307070 (transcriptome)
Hemichordata Enteropneusta	*Ptychodera flava*	Marinegenomics	pfl_40v0_9_20150316_1g2314.t1
		GenBank WGS	LD343027 41534-50098
Hemichordata Enteropneusta	*Ptychodera flava*	GenBank WGS	LD343027 51007-66347
Hemichordata Enteropneusta	*Ptychodera flava*	Marinegenomics	pfl_40v0_9_20150316_1g6997.t1
		GenBank WGS	BCFJ01022326 32811-41459
Hemichordata Enteropneusta	*Saccoglossus kowalevskii*	GenBank proteins	XP_006816582
Hemichordata Enteropneusta	*Saccoglossus kowalevskii*	GenBank proteins	XP_006816581
Hemichordata Enteropneusta	*Saccoglossus kowalevskii*	GenBank proteins	XP_006819810
Hemichordata Enteropneusta	*Schizocardium californicum*	GenBank SRA	SRX1436000
Mollusca Gastropoda Caenogastropoda	*Asolene platae*	AmpuBase	Apl52885
Mollusca Gastropoda Caenogastropoda	*Conus tribblei (1)*	GenBank WGS	LFLW010536118
Mollusca Gastropoda Vetigastropoda	*Haliotis laevigata*	GenBank TSA	GFTT01038064
Mollusca Gastropoda Vetigastropoda	*Haliotis rubra**	GenBank WGS	QXJH01001142
Mollusca Gastropoda Vetigastropoda	*Haliotis rufescens**	GenBank WGS	QGMO01000565
Mollusca Gastropoda Caenogastropoda	*Crepidula novicella*	GenBank TSA	GELE01086894
Mollusca Gastropoda Caenogastropoda	*Lanistes nyassanus*	AmpuBase	Lny24710
Mollusca Gastropoda Caenogastropoda	*Marisa cornuarietes*	AmpuBase	Mco2627

Mollusca Gastropoda Caenogastropoda	*Neverita didyma*	GenBank TSA	GHHQ01002371
Mollusca Gastropoda Caenogastropoda	*Pila ampullacea*	AmpuBase	Pila82769
Mollusca Gastropoda Caenogastropoda	*Pomacea diffusa*	AmpuBase	Pdi16479 (partial)
Mollusca Gastropoda Caenogastropoda	*Pomacea maculata*	AmpuBase	Pma33988 (partial)
Mollusca Gastropoda Caenogastropoda	*Pomacea scalaris*	AmpuBase	Psc4690
Mollusca Gastropoda Caenogastropoda	*Pomacea canaliculata*	GenBank proteins	XP_025109323
		AmpuBase	Pca5338
Mollusca Gastropoda Caenogastropoda	*Rapana venosa*	GenBank TSA	GDIA01047641
Mollusca Gastropoda Caenogastropoda	*Semisulcospira coreana*	GenBank TSA	GGNX01073707
Mollusca Bivalvia Mytiloida	*Limnoperna fortunei (1)*	GenBank Assemby	NFUK01006104
Mollusca Bivalvia Mytiloida	*Bathymodiolus platifrons**	GenBank Assembly	MJUT01033839
Mollusca Bivalvia Mytiloida	*Modiolus philippinarum**	GenBank Assembly	MJUU01021410
Mollusca Bivalvia Mytiloida	*Mytilus galloprovincialis (1)*	GenBank Assembly	APJB011511270
Mollusca Bivalvia Mytiloida	*Mytilus galloprovincialis*	GenBank TSA	GHIK01025031
Mollusca Bivalvia Mytiloida	*Perna canaliculus*	GenBank TSA	GGLA01150624
Mollusca Bivalvia Mytiloida	*Septifer virgatus*	GenBank TSA	GFKS01035611
Mollusca Bivalvia Pterioida	*Pinctada martensi**	GenBank Assembly	CM008066
Mollusca Bivalvia Pterioida	*Pinctada fucata*	Marinegenomics	pfu_aug1.0_4142.1_01638
Mollusca Bivalvia Pterioida	*Pteria penguin*	GeneBank TSA	GEMO01011007
Mollusca Bivalvia Arcoidea	*Scapharca broughtoni*	GenBank TSA	GEXI01046152
Annelida Oligochaeta	*Glossoscolex paulitsus*	GenBank TSA	GBIL01075477
Annelida Polychaeta	*Hydroides elegans**	GenBank Assembly	LQRL01141559
			LQRL01153670
Annelida Polychaeta	*Pygospio elegans*	GenBank TSA	GFPL01035490
Annelida Polychaeta	*Spirobranchus lamarcki*	GenBank TSA	GGGS01192599
UNICELLULAR EUKARYOTES			
Amoebozoa Mycetozoa	*Cavendaria fasciculata*	GenBank proteins	XP_004351949
Amoebozoa Mycetozoa	*Dictyostellium discoideum*	GenBank proteins	XP_640516**
Amoebozoa Mycetozoa	*Polysphondylium pallidum*	GenBank proteins	XP_020429468
Amoebozoa Discosea	*Acanthamoeba castellanii*	GenBank proteins	XP_004368209
Choanoflagellida Salpingoecidae	*Monosiga brevicollis*	GenBank proteins	XP_001742116

Choanoflagellida Salpingoecidae	*Salpingoeca rosetta*	GenBank proteins	XP_004998636
Ciliata	*Ichthyophthirius multifiliis*	GenBank proteins	XP_004027176
Ciliata	*Euplotes focardii*	GenBank proteins	AGU13046**
Ciliata	*Moneuplotes crassus*	GenBank proteins	AGU13047**
Ciliata	*Paramecium tetraurelia*	GenBank proteins	XP_001462315
Ciliata	*Stentor coeruleus*	GenBank proteins	OMJ70617
Ciliata	*Stylonychia lemnae*	GenBank proteins	CDW84776
Ciliata	*Tetrahymena thermophila*	GenBank proteins	XP_001020855**
Heterolobosea	*Naegleria gruberi*	GenBank proteins	XP_002676377
Apusozoa	*Thecamonas trahens*	GenBank proteins	XP_013759080
Oomycetes	*Achlya hypogyna*	GenBank proteins	AIG56379**
Oomycetes	*Saprolegnia diclina*	GenBank proteins	XP_008604251
Oomycetes	*Thraustotheca clavata*	GenBank proteins	AIG55673**

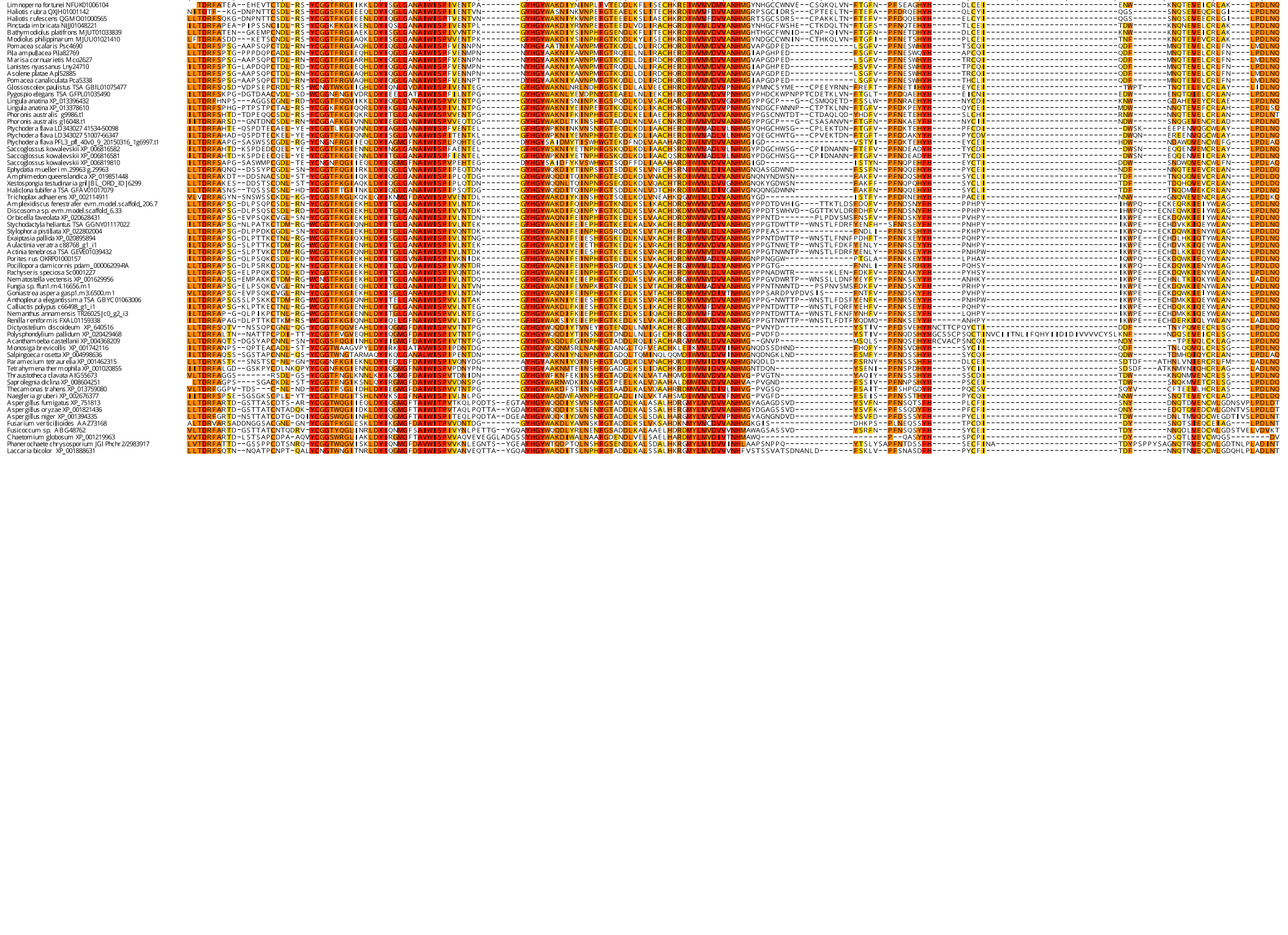
Table S1 : Analysis of neighboring genes for *Lingula anatina*, *Pomacea canaliculata* and *Saccoglossus kowalevskii*, according to the genome browsers at GenBank ; and for *Phoronis australis* and *Ptychodera flava* according to the OIST marine genomics genome browser. Best non-self BLASTP hits against the GenBank Protein database were recorded for each of the three neighboring protein-coding genes on each side of the GH13_1 genes. The BLAST expect values are indicated. The colors indicate putative orthologous genes for duplicated genes. * : The two genes are in tandem. (1) : The tandem *Amy* genes are the last ones on this contig. (°) : No counterpart in *Phoronis australis* (for *L. anatina*) or in *Ptychodera flava* (for *S. kowalevskii*).

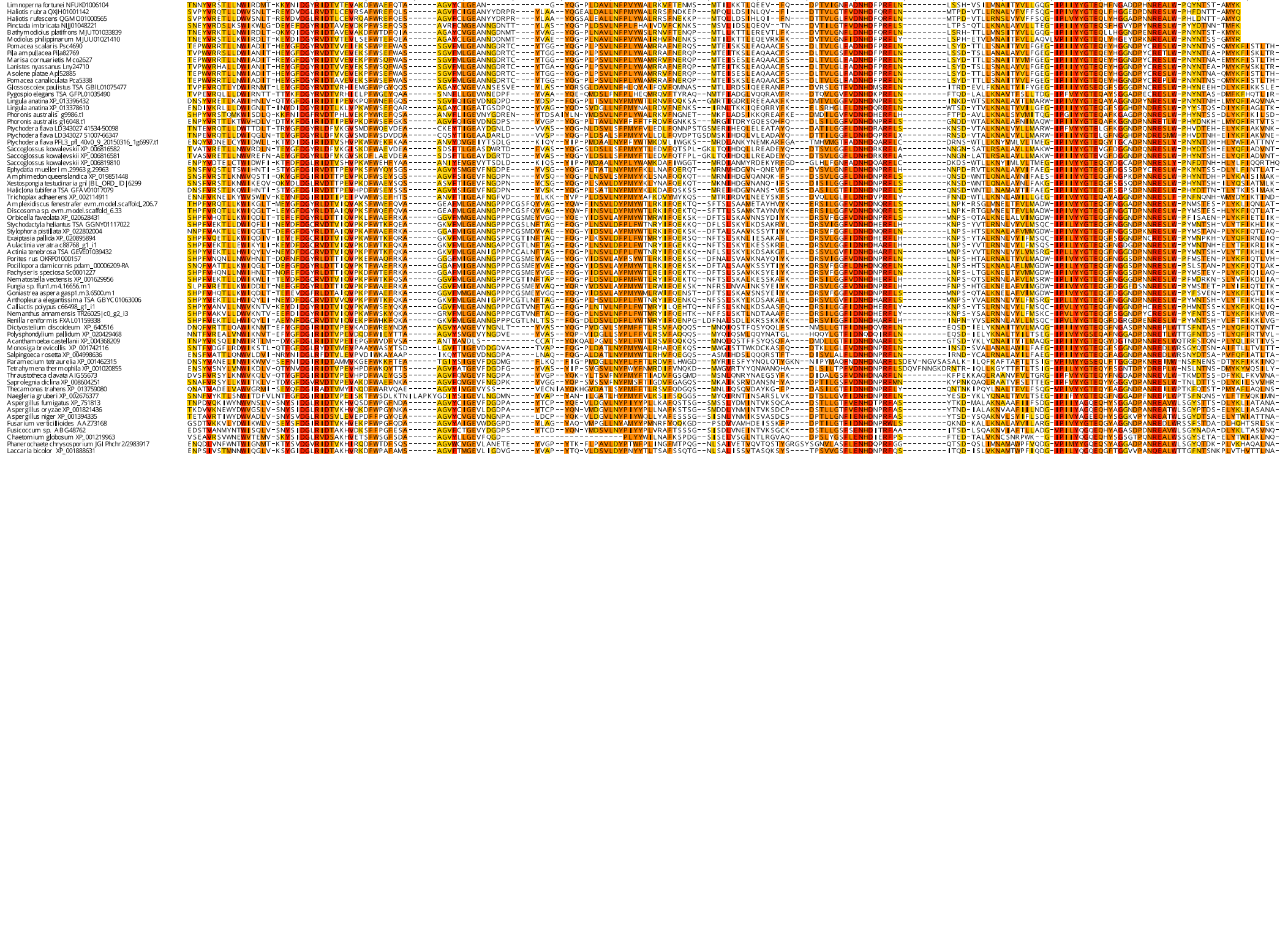
Species/gene	**-3**	**-2**	**-1**	**+1**	**+2**	**+3**
Lingula anatina XP_013396432	XP_013396431 No animal hit (°)	XP_023931706 Vertebrata e-57	XP_013396389 Bivalvia e-11	XP_023931696 Echinodermata e-48	XP_023931697 Hemichordata 0.0	XP_013406295 No animal hit
Lingula anatina XP_013378610	XP_013378627 Annelida e-53	XP_013378628 Bivalvia e-118	XP_013378631 Gastropoda e-59	XP_013378654 Annelida e-51	XP_013378655 Cnidaria e-63	XP_013378611 Cephalochordata e-33
Saccoglossus kowalevskii XP_006816581 XP_006816582*	XP_006816579 No hit (°)	XP_006816580 Hexapoda e-73	XP_006816576 Cephalochordata e-60	(1)	(1)	(1)
Saccoglossus kowalevskii XP_006819810			XP_006819811 No hit (°)	XP_006819812 Placozoa e-47		
Phoronis australis g16048.t1			g16047.t1 Annelida e-71	g16049.t1 No hit	g16050.t1 Bivalvia 0.0	g16051.t1 Brachiopoda e-36
Phoronis australis G9986.t1	g9983.t1 Brachiopoda e-62	g9984.t1 Bivalvia 0.0	g9985.t1 Brachiopoda e-95	g9987.t1 Gastropoda e-77	g9988.t1 Echinodermata e-76	g9989.t1 Brachiopoda 0.0
Ptychodera flava LD343027 41534-66347*		1g2312.t1 No hit	1g2313.t1 Hemichordata e-60	(1)	(1)	(1)
Ptychodera flava 1g6997	1g6994 No hit	1g6995 Hemichordata 0.0	1g6996 Hemichordata e-79	1g6998 No hit	1g6999 Hemichordata e-79	1g7000 Echinodermata e-62

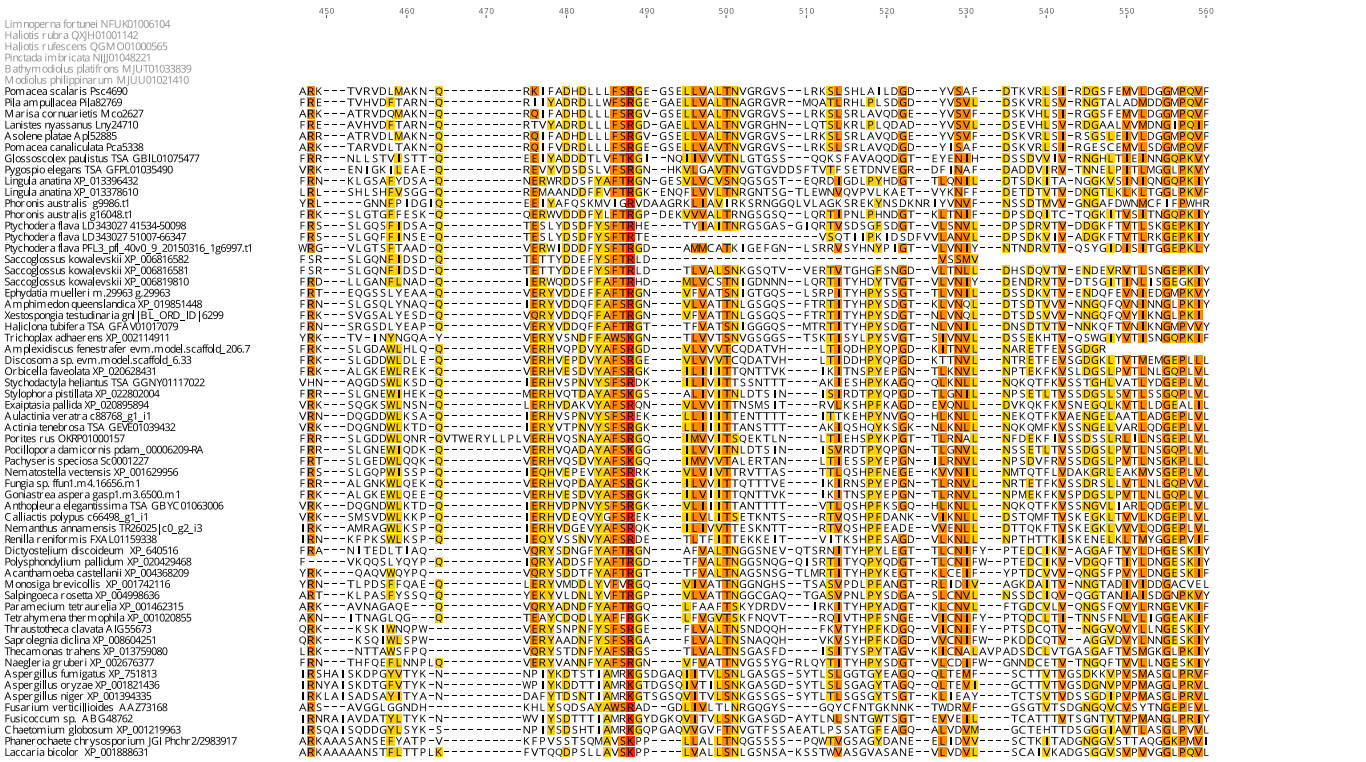
Pomacea canaliculata	XP_025110260 Mollusca e-33	XP_025109347 Annelida e-64	XP_025111249 Mollusca e-91	XP_025109322 Mollusca e-80	XP_025110021 Mollusca e-145	XP_025110160 Mollusca e-103

Table S2 : Results of Tajima's relative rate tests for the comparison of evolutionary speed of *S. kowalevskii* XP_006819810 and its paralog XP_006816581.

outgroup	χ^2	p
A. queenslandica	5.28	0.022
E. muelleri	0.07	0.793
X. testudinaria	8.53	0.003
M. brevicollis	1.61	0.204
S. rosetta	5.43	0.020


Limnoperna fortunei NFUK01006104
Haliotis rubra QXJH01001142
Haliotis rufescens QGMO01000565
Pinctada imbricata NIJJ01048221
Bathymodiolus platifrons MJUT01033839
Modiolus philippinarum MJUU01021410
Pomacea scalaris Psc4690
Pila ampullacea Pila82769
Marisa cornuarietis Mco2627
Lamsites nyassanus Lny24710
Asolene platae Apl52885
Pomacea canaliculata Pca5338
Glossoscolex paulistus TSA GBIL01075477
Pygospio elegans TSA GFPL01035490
Lingula anatina XP_013396432
Lingula anatina XP_013378610
Phoronis australis g9986.t1
Phoronis australis g16048.t1
Ptychodera flava LD343027 415344-50098
Ptychodera flava LD343027 51007-66347
Ptychodera flava PFL3_pfl_40v0_9_20150316_1g6997.t1
Saccoglossus kowalevskii XP_006816582
Saccoglossus kowalevskii XP_006816581
Saccoglossus kowalevskii XP_006819810
Ephydatia muelleri m_29963 g_29963
Amphimedon queenslandica XP_019851448
Xestospongia testudinaria grl|BL_ORD_ID|6299
Haliclona tubifera TSA GFAV01017079
Trichoplax adhaerens XP_002114911
Amplexidiscus fenestrafer evm.model.scaffold_206.7
Discosoma sp. evm.model.scaffold_6.33
Oribicella faveolata XP_020628431
Stychodactyla helianthus TSA GGNY01117022
Stylophora pistillata XP_022802004
Exaiptasia pallida XP_020895894
Aulactinia veratra c88768_g1_i1
Actinia tenebrosa TSA GEVE01039432
Porites rus OKRP01000157
Pocillopora damicornis pdam_00006209-RA
Pachyseris speciosa Sxo001227
Nematostella vectensis XP_001629956
Fungia sp. ffun1.m4.16656.m1
Goniastrea aspera gasp1.m3.6500.m1
Anthopleura elegantissima TSA GBYC01063006
Calliactis polypus c66498_g1_i1
Nemanthus annamensis TR26025|c0_g2_i3
Renilla reniformis FXA10115938
Dictyostelium discoideum XP_640516
Polysphondylium pallidum XP_020429468
Acanthamoeba castellanii XP_004368209
Monosiga brevicollis XP_001742116
Salpingoeca rosetta XP_004998636
Paramecium tetraurelia XP_001462315
Tetrahymena thermophila XP_001020855
Thraustotheca clavata AGS56573
Saprolegnia diclina XP_008604251
Thecamonas trahens XP_013759080
Naegleria gruberi XP_002676377
Aspergillus fumigatus XP_751813
Aspergillus oryzae XP_001821436
Aspergillus niger XP_001394335
Fusarium verticillioides AAZ73168
Fusicoccum sp. ABG48762
Chaetomium globosum XP_001219963
Phanerochaete chrysosporium JGI Phchr 2/2983917
Laccaria bicolor XP_001888631




Limnoperna fortunei NFUK01006104
Haliotis rubra QXJH01001142
Haliotis rufescens QGMO01000565
Pinctada imbricata NIJJ01048221
Bathymodiolus platifrons MJUT01033839
Modiolus philippinarum MJUU01021410
Pomacea scalaris Psc4690
Pila ampullacea Pla82769
Marisa cornuarietis Mco2627
Lamellites nyassanus Lny24710
Asolene platae Apl52885
Pomacea canaliculata Pca5338
Glossoscolex paulistus TSA_GBIL01075477
Pygospio elegans TSA_GFPL01035490
Lingula anatina XP_013396432
Lingula anatina XP_013378610
Phoronis australis_g9986.t1
Phoronis australis_g16048.t1
Ptychodera flava LD343027 41534-50098
Ptychodera flava LD343027 51007-66347
Ptychodera flava PFL3_pfl_40v0_9_20150316_1g6997.t1
Saccoglossus kowalevskii XP_006816582
Saccoglossus kowalevskii XP_006816581
Saccoglossus kowalevskii XP_006819810
Ephydatia muelleri m_25963.g_25963
Amphimedon queenslandica XP_019851448
Xestospongia testudinaria gnl|BL_ORD_ID|6299
Haliclona tubifera TSA_GFAV01017079
Trichoplax adhaerens XP_002114911
Amplexidiscus fenestrafer_evm.model.scaffold_206.7
Discosoma sp._evm.model.scaffold_6.33
Oribicella faveolata XP_020628431
Stychodactyla helianthus TSA_GGNY01117022
Stylophora pistillata XP_022802004
Exaiptasia pallida XP_020895894
Aulactinia veratra c88768_g1_i1
Actinia tenebrosa TSA_GEVE01039432
Porites rus_OKRP01000157
Pocillopora damicornis pdam_00006209-RA
Pachyseris speciosa Scv0001227
Nematostella vectensis XP_001629956
Fungia sp._ffun1.m.4.16656.m1
Goniastrea aspera gasp1.m3.6500.m1
Anthopleura elegantissima TSA_GBYC01063006
Calliactis polypus c66498_g1_i1
Nemanthus annamensis TR26025|c0_g2_i3
Renilla reniformis FXA10115938
Dictyostelium discoideum_XP_640516
Polysphondylium pallidum XP_020429468
Acanthamoeba castellanii XP_004368209
Monosiga brevicollis_XP_001742116
Salpingoeca rosetta XP_004998636
Paramecium tetraurelia XP_001462315
Tetrahymena thermophila XP_001020855
Thraustotheca clavata AGS56673
Saprolegnia diclina XP_008604251
Thecamonas trahens XP_013759080
Naegleria gruberi XP_002676377
Aspergillus fumigatus XP_751813
Aspergillus oryzae XP_001821436
Aspergillus niger XP_001394335
Fusarium verticillioides_AAZ73168
Fusicoccum sp._ABG48762
Chaetomium globosum XP_001219963
Phanerochaete chrysosporium JGI Phchr 2/2983917
Laccaria bicolor_XP_001888631