



HAL
open science

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

Andrea Desiderato, Marcos Barbeitos, Clément Gilbert, Jean-Luc da Lage

► **To cite this version:**

Andrea Desiderato, Marcos Barbeitos, Clément Gilbert, Jean-Luc da Lage. Horizontal transfer and gene loss shaped the evolution of alpha-amylases in bilaterians. *G3*, 2020, 10 (2), pp.709-719. 10.1534/g3.119.400826 . hal-02401035

HAL Id: hal-02401035

<https://hal.science/hal-02401035>

Submitted on 9 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Horizontal transfer and gene loss shaped the evolution of alpha-amylases in** 2 **bilaterians**

3
4
5 Andrea Desiderato*†, Marcos Barbeitos *, Clément Gilbert‡, Jean-Luc Da Lage‡

6
7 *: Graduate Program in Zoology, Zoology Department, Federal University of Paraná, CP
8 19020, Curitiba, Paraná 81531-980, Brazil

9 †: Department of Functional Ecology, Alfred Wegener Institute & Helmholtz Centre for
10 Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

11
12 ‡: Évolution, Génomes, Comportement, Écologie. CNRS, IRD, Université Paris-Sud.
13 Université Paris-Saclay. F-91198 Gif-sur-Yvette, France

14 15 16 17 18 19 **Abstract**

20 The subfamily GH13_1 of alpha-amylases is typical of Fungi, but it is also found in some
21 unicellular eukaryotes (e.g. Amoebozoa, choanoflagellates) and non-bilaterian Metazoa. Since
22 a previous study in 2007, GH13_1 amylases were considered ancestral to the Unikonts,
23 including animals, except Bilateria, such that it was thought to have been lost in the ancestor of
24 this clade. The only alpha-amylases known to be present in Bilateria so far belong to the
25 GH13_15 and 24 subfamilies (commonly called bilaterian alpha-amylases) and were likely
26 acquired by horizontal transfer from a proteobacterium. The taxonomic scope of Eukaryota
27 genomes in databases has been greatly increased ever since 2007. We have surveyed GH13_1
28 sequences in recent data from ca. 1600 bilaterian species, 60 non-bilaterian animals and also in

29 unicellular eukaryotes. As expected, we found a number of those sequences in non-bilaterians:
30 Anthozoa (Cnidaria) and in sponges, confirming the previous observations, but none in
31 jellyfishes and in Ctenophora. Our main and unexpected finding is that such fungal (also called
32 Dictyo-type) amylases were also consistently retrieved in several bilaterian phyla:
33 hemichordates (deuterostomes), brachiopods and related phyla, some molluscs and some
34 annelids (protostomes). We discuss evolutionary hypotheses possibly explaining the scattered
35 distribution of GH13_1 across bilaterians, namely, the retention of the ancestral gene in those
36 phyla only and/or horizontal transfers from non-bilaterian donors.

37

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

40

41 **Introduction**

42 Alpha-amylases are enzymes that are almost ubiquitous in the living world, where they
43 perform the hydrolysis of starch and related polysaccharides into smaller molecules, to supply
44 energy to the organism through digestion. They belong to glycosyl hydrolases, a very large
45 group of enzymes which have been classified in a number of families according to their
46 structures, sequences, catalytic activities and catalytic mechanisms (Henrissat and Davies 1997).
47 Most alpha-amylases are members of the glycoside hydrolase family 13 (GH13), which includes
48 enzymes that can either break down or synthesize α -1,4-, α -1,6- and, less commonly, α -1,2- and
49 α -1,3-glycosidic linkages. Sucrose and trehalose are also substrates for enzymes of this family
50 (MacGregor *et al.* 2001). The numerous family GH13 is divided into 42 subfamilies, of which
51 only three occur in Metazoans: GH13_1, GH13_15 and GH13_24 (Stam *et al.* 2006; Da Lage *et al.*
52 *et al.* 2007; Lombard *et al.* 2014). The latter two include the common animal alpha-amylases, while
53 the former was first described in Fungi for which it represents the canonical alpha-amylase
54 (Stam *et al.* 2006). Da Lage *et al.* (2007) described the subfamilies GH13_15/24 as private to
55 Bilateria among metazoans. In the same article, they retrieved sequences belonging to the
56 subfamily GH13_1 from the sponge *Amphimedon queenslandica* (named *Reniera sp.* in their
57 paper) and the sea anemone *Nematostella vectensis*, besides the unikont choanoflagellates and
58 amoebozoans, and also excavates and ciliates. They dubbed “Dictyo-type” this alpha-amylase,
59 referring to the slime mold *Dictyostelium discoideum* (Amoebozoa Mycetozoa). The authors
60 proposed that this amylase, ancestral to the Unikont clade, is shared among non-bilaterian
61 metazoans (e.g. sponges, sea anemones and corals, and Placozoa), but was replaced in Bilateria
62 by an alpha-amylase of bacterial origin, whose sequence is close to the typical animal amylases.

63 Given that a wealth of new genomes have been sequenced in the twelve years after that
64 publication, we decided to explore again the diversification of this enzyme subfamily among the
65 Eukaryota. We will focus mainly on Metazoa, in which we show unexpected situations of co-
66 occurrence of both subfamilies GH13_1 and GH13_15/24 in the same genomes. We will discuss
67 two mutually exclusive explanations that may be proposed: either the retention of the ancestral
68 GH13_1 gene along with the typical bilaterian GH13_15/24 in some phyla, or horizontal
69 transfer(s) from non-bilaterian animal donor(s) which would have to be identified.

70

71

72 **Materials and methods**

73 In order to further characterize the distribution of GH13_1 genes in Metazoa, we used
74 the sequence of the sponge *Amphimedon queenslandica* GH13_1 (GenBank XP_019851448) as
75 a query to perform BLASTP and TBLASTN searches on various online databases available in
76 Genbank (nr, proteins, genomes, assembly, SRA, TSA, WGS), and also in the more specialized
77 databases compagen.org, marinegenomics.oist.jp, reefgenomics.org, marimba.obs-vlfr.fr,
78 vectorbase.org, PdumBase (pdumbase.gdcb.iastate.edu), AmpuBase
79 (<https://www.comp.hkbu.edu.hk/~db/AmpuBase/index.php>) (Ip *et al.* 2018), between October
80 2018 and August 2019. Fungi were not searched further in this study because they are known to
81 have a GH13_1 member as the usual alpha-amylase. To increase the chances to retrieve potential
82 cnidarian or ctenophoran sequences, the starlet sea anemone *Nematostella vectensis* amylase
83 (XP_001629956) was also used to query those databases. After the discovery of GH13_1-like
84 sequences in Bilateria, the sequence XP_013396432 of the brachiopod *Lingula anatina* was also
85 used for specific search in Bilateria. Non-animal eukaryote species were investigated using the
86 *Dictyostelium discoideum* sequence XP_640516 as query. We chose a stringent BLAST
87 threshold because glycosyl hydrolases from other GH13 subfamilies might be retrieved
88 otherwise, owing to the presence of stretches of amino acids that are conserved across the
89 subfamilies despite other enzymatic specificities (Janeček 1994; Janeček *et al.* 2014; MacGregor
90 *et al.* 2001; van der Kaaij *et al.* 2007). Therefore, the BLAST hits (or High-scoring segment
91 pairs HSPs) were considered further when expectation values (e-values) were better (lower) than
92 1e-100 for BLASTP or 1e-75 for TBLASTN in annotated or assembled genomes or
93 transcriptomes which returned full-size or near full-size GH13_1 sequences. When only partial
94 sequences could be retrieved using BLAST we collected a large genome region encompassing
95 and flanking the BLAST hit and tried to reconstitute the full-size sequence. The stringent
96 threshold was obviously not applied to constitutively small HSPs such as SRA (sequence read
97 archives). These were only considered when several highly significant hits (typically 1e-10)
98 covered a large part of the query sequence. Since SRA HSPs generally did not overlap, we could
99 not assemble longer sequences and thus we did not use such sequences in alignments or
100 phylogenies. SRA and transcriptome databases are prone to contamination, thus we checked by
101 reciprocal BLAST that the retrieved sequences were not contaminations. SRA databases were
102 used firstly when no or few assembled genomes or transcriptomes were available (e.g. Nemertea,
103 Bryozoa). If a GH13_1 sequence was found in an annotated or assembled genome, SRA HSPs
104 were also used in order to increase the sampling of related taxa and then added some support to
105 the presence of GH13_1 in the lineage considered. On the other hand, we considered that within
106 a given lineage, the absence of GH13_1 sequence in a reliable annotated or assembled genome,

107 combined to the detection of GH13_1 in SRA databases from related species would suggest that
108 the GH13_1 gene was lost within this lineage in some taxa but not all. Conversely, if no GH13_1
109 sequence at all was found in any annotated genome and in any other database, we considered that
110 the gene was lost in an ancestor of this lineage. When working with “assembled genomes” (non-
111 annotated), we reconstituted exon-intron gene structure as well as the the protein sequence from
112 the TBLASTN results. Finally, for phylogenetic analyses we kept only sequences which lay
113 inside long contigs, or full-size or near full-size transcripts. We also checked once again the
114 absence of animal-type alpha-amylase (GH13_15 or 24) outside the Bilateria using the sequence
115 of the bivalve *Corbicula fluminea* (AAO17927) as a BLASTP query. The CAZy database
116 (cazy.org (Lombard *et al.* 2014)), which is devoted to glycosyl hydrolases and related enzymes
117 was used to check assignment of some of the sequences we found to the GH13_1 subfamily.

118 Intron-exon gene structures were recovered either from alignments between genomic
119 sequences and their mRNA counterparts, or using annotated graphic views when available in the
120 databases. In cases of likely erroneous annotations we reanalyzed the gene region by eye,
121 correcting dubious frameshifts if necessary (see. Fig. S1 as an example). In some cases, for
122 unannotated genes, the N-terminal and/or the C-terminal parts of the retrieved genomic
123 sequences were uncertain, and were not retained in the analyses.

124 Alignments were performed using MUSCLE (Edgar 2004), as implemented in Geneious
125 (Biomatters Ltd.). A maximum likelihood (ML) tree was built using PhyML’s (Guindon and
126 Gascuel 2003) current implementation at the phylogeny.fr portal (Dereeper *et al.* 2008). To this
127 end, we first trimmed the N-terminal protein sequences up to the first well conserved motif
128 LLTDR. C-terminal parts were also truncated at the last well aligned stretch. Gaps were removed
129 from the alignments and data were analyzed under WAG (Whelan and Goldman 2001) with
130 among-site rate variation modeled by four discretized rate categories sampled from a gamma
131 distribution. Both the alpha parameter and the proportion of invariable sites were estimated from
132 the data. The robustness of the nodes was estimated using an approximate likelihood ratio test
133 (aLRT) (Anisimova and Gascuel 2006). The tree was drawn at the iTOL website (Letunic and
134 Bork 2016). Metazoans and choanoflagellates were clustered as the ingroup.

135 The protein sequences, Fasta alignment and Newick-formatted tree are available at FigShare:
136 https://figshare.com/articles/GH13_1_metazoa/9959369. Supplementary Figures and Tables are
137 available at FigShare: https://figshare.com/articles/Suppl_data_GH13_1/9975956

138 **Results**

139 The sequences retrieved from the databases are listed in Table 1. The metazoans investigated are
140 listed in Tables S1 (non-bilaterians) and S2 (bilaterians) with indication of the current state of
141 genome/transcriptome sequencing, the database, the presence or absence of GH13_1 sequences,
142 and the number of gene copies, where possible. A general protein alignment of the sequences
143 found in this study along with already known GH13_1 sequences is shown in Fig. S2.

144

145 *GH13_1 sequences retrieved from unicellular taxa*

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

156

157 *GH13_1 sequences retrieved from non-bilaterian animals*

158 In Cnidaria, a number of GH13_1 sequences were recovered from many Anthozoa
159 species (sea anemones, corals and allies), from genome as well as transcriptome data, at the
160 Reefgenomics database (Table S1). Interestingly, we found no alpha-amylase sequences at all in
161 Medusozoa (jellyfishes, hydras) nor in Endocnidozoa (parasitic cnidarians). In the general tree
162 (Fig. 1), cnidarian sequences form a clear cluster with two main branches, grouping Actiniaria
163 (sea anemones) and Pennatulacea (soft corals) on one branch, and Scleractinia (hard corals) and
164 Corallimorpharia (mushroom anemones) on the other branch.

165 In sponges (Porifera), data were less abundant. No alpha-amylase sequence was found
166 in *Sycon ciliatum* (Calcarea) and *Oscarella carmela* (Homoscleromorpha). All the sequences we
167 retrieved belonged to Demospongiae. Similarly, we found no amylase sequence at all in the
168 phylum Ctenophora (*Mnemiopsis leidyi*, *Pleurobrachia bachei*), the phylogenetic position of
169 which is controversial: it has been recovered as the most basal metazoan (Whelan *et al.* 2017), as

170 Cnidaria's sister group e.g. Simion *et al.* 2017, Philippe *et al.* 2009), re-establishing
171 Coelenterata, and also as the earliest branch in the Eumetazoa (animals with a digestive cavity
172 and/or extra cellular digestion) e.g. Pisani *et al.* 2015.

173

174 *GH13_1 sequences retrieved from bilaterian animals*

175 The surprising finding of this study, on which we will focus our attention, is the
176 consistent, albeit sparse, occurrence of GH13_1 alpha-amylase sequences in several bilaterian
177 phyla: hemichordates, which are deuterostomes, brachiopods, phoronids (Brachiozoa) and
178 Bryozoa, and in some molluscs and annelids (Eutrochozoa), which are all protostomes. In the
179 well annotated genomes of the brachiopod *Lingula anatina* and the phoronid *Phoronis australis*,
180 two paralogs were found (Table 1). In both species, the two copies are located on different
181 contigs. The paralog sequences are rather divergent, given their positions in the tree (Fig 1) and
182 each paralog groups the two species together. This indicates that not only duplication, but also
183 the divergence between paralogs is ancestral to these species, dating back at least to basal
184 Cambrian, according to the TimeTree database (Kumar *et al.* 2017). GH13_1 sequences were
185 found in other brachiopods and phoronids as sequence reads (SRA) from transcriptome data
186 only, with no available genomic support (listed in Table 1 and S2). We must be cautious when
187 only transcriptome data are available, as transcripts from contaminating symbionts or parasites
188 may generate false positives (Borner and Burmester 2017) and/or the lack of expression of the
189 targeted sequence in the investigated tissues may lead to false negatives. However, seven
190 different brachiopod species returned positive hits, giving some robustness to our finding. Two
191 phyla are related to Brachiozoa : Bryozoa and Nemertea (Kocot 2015; Luo *et al.* 2018, but see
192 Marlétaz *et al.* 2019). We found clues for the presence of GH13_1 in four Bryozoa species, but
193 only transcriptome reads were available. In contrast, in Nemertea, none of the 14 species
194 investigated returned any GH13_1 sequence, including the annotated genome of *Notosperma*
195 *geniculatus*.

196 Similarly, we found three gene copies in the genomes of the hemichordates
197 *Saccoglossus kowalevskii* and *Ptychodera flava*. In both species, two copies are close to each
198 other (XP_006816581 and XP_006816582 in *S. kowalevskii*, and their counterparts in *P. flava*)
199 as shown by the topology of the gene tree (Fig. 1). This could suggest independent gene
200 duplication in each species. However, we observed that the two duplicates were arranged in
201 tandem in both species, which would rather suggest concerted evolution of two shared copies. In

202 *P. flava*, this genome region is erroneously annotated as a single gene at the OIST Marine
203 Genomics database. The third paralog is very divergent from the two other copies, so its
204 divergence from the ancestral copy probably occurred before the species split, as well. The three
205 copies were therefore probably already present before the split of the two lineages, some 435
206 mya (Kumar *et al.* 2017). Three other hemichordate species, *Schizocardium californicum*,
207 *Torquaratoridae antarctica* and *Rhabdopleura sp.* harbor a GH13_1 gene, as shown by SRA
208 search in GenBank (Table 1). A positive result was also retrieved from the genome of
209 *Glandiceps talaboti* (Héctor Escrivà, Oceanology Observatory at Banyuls-sur-mer, personal
210 communication).

211 In molluscs, we found BLAST hits with significant e-values in gastropod species from
212 two clades only, the Vetigastropoda (e.g. the abalone *Haliotis sp.*) and the Caenogastropoda (e.g.
213 Ampullariidae such as *Pomacea canaliculata*). We consistently found one copy in eight species
214 belonging to the family Ampullariidae. In *P. canaliculata*, the genome of which has been well
215 annotated, the GH13_1 sequence (XP_025109323) lies well inside a 26 Mbp long scaffold
216 (linkage group 10, NC_037599) and is surrounded by *bona fide* molluscan genes (Table S3).
217 GH13_1 sequences were found in other Caenogastropoda from SRA or transcriptome databases
218 (Table 1 and S2). We also found GH13_1 sequences in several bivalve clades: Mytiloidea (e.g.
219 the mussel *Mytilus galloprovincialis*), Pterioidea (e.g. the pearl oyster *Pinctada imbricata*),
220 Arcoidea (e.g. *Scapharca broughtoni*) and in the Unionoidea *Cristaria plicata*. For sequences
221 retrieved from the TSA or SRA databases (see Table 1), whose issues were mentioned above, we
222 performed reciprocal BLAST in GenBank nr. Almost always *Lingula anatina* was recovered as
223 the best hit. However, as an example of the necessary careful examination of results, we found a
224 significant HSP in a transcriptome database of the sea hare *Aplysia californica* (TSA
225 GBDA01069500). This sequence was not found in the well annotated *A. californica* genome,
226 and turned out to be related to ciliates. We found no occurrence of GH13_1 in Veneroidea,
227 Pectinoidea and Ostreoidea, for which annotated and/or assembled genomes exist, nor in
228 cephalopods.

229 In annelids, we found occurrences of GH13_1 genes in a few species, the genomes of which are
230 still not fully assembled, namely the “polychaetes” *Hydroides elegans*, *Pygospio elegans* and
231 *Spirobranchus lamarcki* but not in the well-annotated genome of *Capitella teleta*. We also
232 recovered HSPs from the clitellate *Glossoscolex paulistus* but not from *Amyntas corticis* or
233 *Eisenia fetida*. We found no GH13_1 sequences in Hirudinea (leeches). To summarize, in
234 molluscs as well as in annelids, the presence of GH13_1 genes is scattered and patchy across and

235 within lineages. Interestingly, we found that some of the mollusc GH13_1-like sequences,
236 especially in bivalves, were much shorter, either truncated at the C-terminal, or this region was
237 so divergent from the query sequence (*L. anatina*) that it was impossible to identify, assemble
238 and align it with our data set (Fig. S2). In addition, we found that the annelid *Hydroides elegans*
239 had an internal deletion, which precluded its inclusion in the phylogenetic analysis. This suggests
240 that those sequences may not have alpha-amylase activity.

241

242 *Gene tree analysis: position of bilaterian sequences*

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

253 All Cnidarian orthologs form a well-supported cluster. The sister relationship between
254 Corallimorpharia and Scleractinia reflects what was recovered in species trees using different
255 markers (e.g. Rodríguez *et al.* 2014), although the Scleractinia topology disagrees with previous
256 phylogenetic analyses of the order (e.g. Barbeitos *et al.* 2010). The other cluster within Cnidaria
257 is mainly composed of actinarian (sea anemones) sequences, but it also includes, with strong
258 support, the sequence queried from the sea pen *Renilla reniformis* (order Pennatulacea). This
259 order belongs to the sub-class Octocorallia and not to Hexacorallia, the monophyletic sub-class
260 in which scleractinians, corallimorpharians and sea anemones are found (e.g. Rodríguez *et al.*
261 2014). We used RAXML-NG v0.80 (Kozlov *et al.* 2019) to conduct a constrained search under
262 WAG for a ML tree in which Hexacorallia was monophyletic and *R. reniformis* was placed as its
263 sister group (e.g. Chang *et al.* 2015; Zapata *et al.* 2015)) and employed a simple LR test to
264 statistically evaluate the difference between the observed and expected (phylogenetic)
265 placement of the *R. reniformis* sequence (Kozlov *et al.* 2019). The log-likelihood difference
266 between the unconstrained (lnLh = -29,155.37) and constrained (lnLh = -29,208.38) ML tree

267 scores was 53.01. According to Kass and Raftery (1995), there is very strong support for the
268 highest likelihood hypothesis (in our case, the ML tree in Fig. 1) when the double of this
269 difference (i.e. $2 \times 53.01 = 106.02$) exceeds 10 log-likelihood units. Thus, there is significant
270 inconsistency between the position of *R. reniformis*' GH13_1 copy and the phylogenetic
271 placement of this species. This may be due to a horizontal transfer event that would have
272 occurred within Cnidaria, but additional data from well-sequenced Pennatulacea would be
273 welcome to check this possibility. Nevertheless, it is noteworthy that the genome of
274 *Dendronephthya gigantea* (Octocorallia, order Alcyonacea) returned no result. Most bilaterian
275 sequences are clustered with Cnidaria, as phylogenetically expected in the case of a shared
276 ancestral gene, as a robust cluster grouping one Brachiozoa (brachiopod/phoronid) copy, the
277 molluscs and the annelids, which is consistent with the phylogeny. However, the tandem
278 hemichordate duplicates and the other Brachiozoa genes are not included in the bilaterian clade,
279 but remain ingroup relative to the sponge sequences.

280 Interestingly, the two remaining hemichordate sequences are the earliest diverging lineage of the
281 Metazoa + Choanoflagellata cluster, since they are branched with the placozoan *Trichoplax*
282 *adhaerens* sequence, this relationship being strongly supported whatever the tree reconstruction
283 method employed (Fig. 1, and data not shown). In order to check for the possibility of a long
284 branch attraction (LBA), which would artificially cluster hemichordate and placozoan sequences,
285 we performed Tajima's relative rate tests (Tajima 1993) using MEGA7 (Kumar *et al.* 2016). The
286 sequence of *S. kowalevskii* XP_006819810, suspected to evolve fast, was compared with its
287 paralog XP_006816581, using five different outgroups, i.e. the three sponges and the two
288 choanoflagellates. Unexpectedly, the χ^2 tests returned non-significant values in two tests and
289 significant values in three tests (Table S4). Therefore, with our data, LBA cannot be entirely
290 ruled out in this particular case.

291

292 *Analysis of intron positions*

293 Intron positions may be valuable markers when reconstituting gene histories. We
294 identified 56 intron positions from the subset of species of the general tree for which we could
295 find data (Fig. 2). Only one intron position is widely shared among these GH13_1 gene
296 sequences. It is the first position reported in the alignment, and it lies just upstream to the first
297 conserved part of the alignment. The main observation is the numerous conserved positions
298 across bilaterian sequences (10 positions), and between bilaterian sequences and the sponge and

299 the Placozoa (7 positions). In addition, three positions are common to bilaterians and the
300 choanoflagellate *Monosiga brevicollis*. In contrast, the Cnidaria have few introns, with positions
301 different from the sponge and the bilaterians, except for position 1. The other species under
302 examination, i.e. protists and fungi, have essentially specific intron positions. Therefore, the
303 overall conservation of intron positions across bilaterians + sponges is a further argument to state
304 that an explanation of the occurrence of GH13_1 alpha-amylases in some bilaterians does not
305 involve non-animal species.

306

307

308 **Discussion**

309 The evolutionary scenario proposed by Da Lage *et al.* 2007, suggested that the GH13_1
310 alpha-amylase gene ancestral to Unikonts (Amoebozoa and Opisthokonts, i.e. Fungi and
311 Metazoa/Choanoflagellata) was totally absent from Bilateria, due to its complete replacement by
312 a new alpha-amylase, originating from a bacterium through HGT. Here, we have shown that a
313 limited number of bilaterian lineages, all aquatic species, namely hemichordates, brachiozoans,
314 bryozoans, and some sparse molluscs and annelids, actually do harbor GH13_1 alpha-amylase
315 genes. Note that all those species also have at least one classical animal alpha-amylase of the
316 GH13_15/24 subfamilies. Several species with whole genome well sequenced and annotated
317 were found to harbor such genes in each phylum Hemichordata, Brachiozoa and molluscs. They
318 were investigated in more details, especially regarding the genomic environment of their
319 GH13_1 genes. We are quite confident that the GH13_1 sequences we found are not due to
320 contaminating DNA. First, the bilaterian sequences retrieved from annotated genomes were
321 inside long contigs, and mostly surrounded by genes showing bilaterian best BLAST hits (Table
322 S3). However, the *S. kowalevskii* XP_006819810 gene could appear somewhat dubious, since it
323 is placed at the distal end of a contig, with only two other genes on the contig (Table S3), one of
324 which has a placozoan best hit. But its *P. flava* counterpart is well inside a gene-rich contig.
325 Therefore, these seemingly non-bilaterian genes are well in bilaterian genomic contexts. Second,
326 a lot of additional sequences from other species belonging to these phyla were gathered from
327 more sketchy data, i.e. lower-quality assembled genomes, transcriptomes or sequence read
328 archive databases, which added some support to the presence of these amylase genes. Although
329 transcriptome and rough genomic data should be handled with care, this lends support to our
330 observations. Moreover, reciprocal BLAST from the transcriptome hits always returned a

331 bilaterian (*L. anatina* or *S. kowalevskii*) best hit, not fungal, protist or other non-bilaterian
332 GH13_1 sequence.

333 The new data unveils an evolutionary story more complicated than previously supposed. There
334 are two alternative explanations. The first explanation is that several HGTs occurred from non-
335 bilaterian to both hemichordate and Lophotrochozoa ancestors. The second explanation is that
336 the ancestral GH13_1 gene was not lost in all bilaterian lineages, but remained (given the current
337 data) in hemichordates, Brachiozoa, Bryozoa, and in scattered lineages across Mollusca and
338 Annelida.

339 The hypothesis of HGT requires several such events between metazoans. It implies that
340 HGTs obviously happened after the split of the two main branches of bilaterians, protostomes
341 and deuterostomes, otherwise the transferred copies should have been lost in most phyla, like in
342 the alternative hypothesis. More precisely, in the case of Lophotrochozoa, this would have
343 occurred before the diversification of this clade and after its divergence from the Platyzoa, some
344 700 mya (Kumar *et al.* 2017); in the case of hemichordates, after diverging from their common
345 ancestor with the echinoderms, and before the divergence between *S. kowalevskii* and
346 *Ptychodera flava*, i.e. between 657 and ca. 435 mya (Kumar *et al.* 2017). Therefore, we may
347 infer *at least* two HGTs, each early in the evolution of the phyla, with a number of subsequent
348 losses in Lophotrochozoa (Fig. 3A). The donor species, given the sequence clustering in the
349 trees, could be related to cnidarians. However, we have underlined that the intron-exon structures
350 of the bilaterian sequences were most similar to the one of the sponge, and that the cnidarian
351 GH13_1 amylases had very different structures. This may be possible if the donors were related
352 to cnidarians, perhaps an extinct phylum or an ancestor of extant Cnidaria, but had conserved the
353 ancestral structures exemplified by the sponge and the placozoan. Indeed, if the structure shared
354 by the sponge, the placozoan and the bilaterians reflects the ancestral state, cnidarians must have
355 undergone a drastic rearrangement of the intron-exon structure of this gene. This would be in
356 line with the long internal branch leading to this clade in the trees (Fig. 1), which suggests
357 accelerated evolution.

358 The alternative hypothesis of massive GH13_1 gene loss in most phyla except the ones
359 where we found such sequences seems no more parsimonious. It requires many losses, the
360 number of which depends on the phylogeny used. For instance, considering the phylogeny
361 shown in Fig. 3B, regarding deuterostomes, one loss occurred in echinoderms and another one in
362 chordates. In protostomes, one GH13_1 loss in ecdysozoans, and independent losses in Platyzoa
363 and in several lophotrochozoan lineages would be required to produce the observed pattern.

364 However, although not parsimonious in terms of number of events, we would rather
365 favor the gene loss hypothesis, because this is a common phenomenon, especially given how
366 ubiquitous co-option is (Flores and Livingstone 2017; Hejnlol and Martindale 2008). In this
367 respect, the GH13_15/24 gene that was acquired from a bacterium is a type of horizontal transfer
368 akin to what Husnik and McCutcheon called a “maintenance transfer” since it allowed the
369 original function to be maintained while the primitive GH13_1 gene became free to evolve or
370 even to be lost (Husnik and McCutcheon 2018) (see also Da Lage *et al.* 2013). In contrast, while
371 numerous cases of HGT from bacteria to metazoans, or from fungi to metazoans have been
372 reported (e.g. Wybouv *et al.* 2016; Dunning Hotopp 2011, 2018; Haegeman *et al.* 2011; Crisp *et*
373 *al.* 2015; Cordaux and Gilbert 2017), very few HGT events have been inferred that involve a
374 metazoan donor and a metazoan receiver (Rödelsperger and Sommer 2011; Graham *et al.* 2012;
375 Gasmi *et al.* 2015). Thus, our current knowledge on HGT suggests that this type of transfer
376 might be very rare between metazoans, and that two or more such events would be quite unlikely
377 to explain the current taxonomic distribution of metazoan GH13_1 genes. In addition, it has been
378 shown that a seemingly patchy gene distribution suggestive of HGT may, after more
379 comprehensive taxon sampling, turn out to be rather due to recurrent gene losses, as discussed in
380 Husnik and McCutcheon (2018). The conservation of the intron-exon structure across phyla,
381 probably ancestral to the metazoans, would not be surprising (Sullivan *et al.* 2006 ; Srivastava *et*
382 *al.* 2010 ; Srivastava *et al.* 2008). For instance, 82% of human introns have orthologous introns
383 in *T. adhaerens* (Srivastava *et al.* 2008).

384

385 In the present study we used the results of BLAST searches (BLASTP and TBLASTN)
386 as raw material using the GH13_1-like alpha-amylases found in non-bilaterian animals (Da Lage
387 *et al.* 2007) as query sequences. The stringent threshold we have set avoids retrieving irrelevant
388 sequences belonging to other GH13 subfamilies or even other GH families. For instance, HMM
389 search, such as in PFAM (pfam.xfam.org), shows that the domain composition of e.g. the
390 *Lingula anatina* sequence XP_013396432 consists in an alpha-amylase domain linked to a
391 DUFF1966 domain (DUFF1266 is also present in several fungal proteins, including obviously
392 the GH13_1 amylase). The alpha-amylase domain is actually present in many glycosyl hydrolase
393 families. Interestingly, the sequences found in some molluscs do not have a complete alpha-
394 amylase domain, because they are shorter than usual (see Results). We assumed nonetheless that
395 all the sequences we recovered belong to the GH13_1 subfamily, due to sequence similarities, as
396 shown by the easy sequence alignment. Further, some of them have been assigned to this

397 subfamily in the reference database CAZy.org (see Table 1). In addition, if we add sequences
398 from the closest subfamilies, namely GH13_2 or GH13_19 (Stam *et al.* 2006) in the alignment
399 and in the phylogenetic tree, the putative GH13_1 and the ascertained GH13_1 remain well
400 clustered together (not shown). It is possible that modifications of a few amino acid positions
401 could bring a change in the substrate or catalytic activity. For instance, concerning the substrate
402 affinity, when the genome of *L. anatina* was released, the authors hypothesized a
403 biomineralization pathway that involves acid proteins, as found in scleractinians and molluscs
404 (Marin *et al.* 2007; Ramos-Silva *et al.* 2013). Given the calcium binding activity of alpha-
405 amylases (Boel *et al.* 1990; Grossman and James 1993; Svensson 1994; Pujadas and Palau
406 2001), the presence of both GH13_1 and GH13_15/24 subfamilies in *L. anatina* opens the
407 possibility for the neofunctionalization of one of them in the biomineralization process. In the
408 analyses performed by those authors, no amylase was found in the shell matrix, but this does not
409 exclude the possibility of its presence in the pathway. Moreover, the fact that in some molluscs,
410 the sequences are incomplete compared to the brachiopod query or to the sponge and cnidarian
411 GH13_1 amylases, and therefore probably devoid of an amylolytic function, would add credence
412 to another function, especially considering that they are transcribed. This conjecture requires
413 further investigation. On the other hand, the full-size GH13_1 sequences only present in a few
414 bilaterians could have remained true alpha-amylases with the classical function, but this would
415 make even more enigmatic why they have been conserved, either by descent or by horizontal
416 transfer.

417

418

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

426

427 **References :**

428 Anisimova, M., and O. Gascuel, 2006 Approximate likelihood-ratio test for branches: A fast,
429 accurate, and powerful alternative. *Systematic Biology* 55 (4):539-552.

430 Barbeitos, M.S., S.L. Romano, and H.R. Lasker, 2010 Repeated loss of coloniality and
431 symbiosis in scleractinian corals. *Proc Natl Acad Sci U S A* 107 (26):11877-11882.

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

435 Borner, J., and T. Burmester, 2017 Parasite infection of public databases: a data mining
436 approach to identify apicomplexan contaminations in animal genome and transcriptome
437 assemblies. *BMC Genomics* 18 (1):100.

438 Chang, E.S., M. Neuhof, N.D. Rubinstein, A. Diamant, H. Philippe *et al.*, 2015 Genomic
439 Insights into the Evolutionary Origin of Myxozoa within Cnidaria. *Proceedings of the*
440 *National Academy of Sciences of the U.S.A.* 112 (48):14912–14917.

441 Cordaux, R., and C. Gilbert, 2017 Evolutionary significance of *Wolbachia*-to-animal horizontal
442 gene transfer: Female sex determination and the f element in the isopod *Armadillidium*
443 *vulgare*. *Genes* 8 (7): doi: 10.3390/genes8070186.

444 Crisp, A., C. Boschetti, M. Perry, A. Tunnacliffe, and G. Micklem, 2015 Expression of
445 multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate
446 genomes. *Genome Biology* 16:50

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

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

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

455 Dunning Hotopp, J.C., 2011 Horizontal gene transfer between bacteria and animals. *Trends in*
456 *Genetics* 27 (4): 157-163.

457 Dunning Hotopp, J.C., 2018 Grafting or pruning in the animal tree: lateral gene transfer and
458 gene loss? *BMC Genomics* 19:470.

459 Edgar, R.C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high
460 throughput. *Nucleic Acids Research* 32 (5):1792-1797.

461 Flores, R.L., and B.T. Livingstone, 2017 The skeletal proteome of the sea star *Patiria miniata*
462 and evolution of biomineralization in echinoderms. *BMC Evolutionary Biology* 17
463 (125):1-14.

464 Gasmi, L., H. Boulain, A. Hua-Van, K. Musset, A.K. Jakubowska *et al.*, 2015 Recurrent
465 domestication by Lepidoptera of genes from their parasites mediated by Bracoviruses.
466 *PLoS Genetics* 11 (9):e1005470.

467 Graham, L.A., J. Li, W.S. Davidson, and P.L. Davies, 2012 Smelt was the likely beneficiary of
468 an antifreeze gene laterally transferred between fishes. *BMC Evolutionary Biology* 12
469 (190): doi: 10.1186/1471-2148-12-190.

470 Grossman, G.L., and A.A. James, 1993 The salivary glands of the vector mosquito *Aedes*
471 *aegypti*, express a novel member of the amylase gene family. *Insect Molecular Biology*
472 1 (4):223-232.

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

475 Haegeman, A., J.T. Jones, and E.G. Danchin, 2011 Horizontal gene transfer in nematodes: a
476 catalyst for plant parasitism? *Molecular Plant and Microbe Interactions* 24 (8):879-887.

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

479 Henrissat, B., and G. Davies, 1997 Structural and sequence-based classification of glycoside
480 hydrolases. *Current Opinion in Structural Biology* 7 (5):637-644.

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

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

485 Janeček, Š., 1994 Sequence similarities and evolutionary relationships of microbial, plant and
486 animal alpha-amylases. *European Journal of Biochemistry* 224:519-524.

487 Janeček, Š., B. Svensson, and E.A. MacGregor, 2014 α -amylase: an enzyme specificity found
488 in various families of glycoside hydrolases. *Cellular and Molecular Life Science*
489 71:1149-1170.

490 Kass, R.E., and A.E. Raftery, 1995 Bayes Factors. *Journal of the American Statistical*
491 *Association* 90 (430):773–795.

492 Kocot, K.M., 2015 On 20 years of Lophotrochozoa. *Organisms Diversity and Evolution*
493 16:329-343.

494 Kocot, K.M., T.H. Struck, J. Merkel, D.S. Waits, C. Todt *et al.*, 2017 Phylogenomics of
495 Lophotrochozoa with considerations of systematic error. *Systematic Biology* 66
496 (2):256-282.

497 Kozlov, A.M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis, 2019 RAxML-NG: A Fast,
498 Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference.
499 *Bioinformatics* doi.org/10.1093/bioinformatics/btz305.

500 Kumar, S., G. Stecher, M. Suleski, and S.B. Hedges, 2017 TimeTree: a resource for timelines,
501 timetrees, and divergence times. *Molecular Biology and Evolution* 34 (7):1812-1819.

502 Kumar, S., G. Stecher, and K. Tamura, 2016 MEGA7: Molecular Evolutionary Genetics
503 Analysis Version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-
504 1874.

505 Laumer, C.E., H. Gruber-Vodicka, M.G. Hadfield, V.B. Pearse, A. Riesgo *et al.*, 2018 Support
506 for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. *eLife*
507 7:e36278.

508 Letunic, I., and P. Bork, 2016 Interactive tree of life(iTOL) v3: an online tool for the display
509 and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44 (W1):W242-
510 245.

511 Lombard, V., H. HGolaconda Ramulu, E. Drula, P.M. Coutinho, and B. Henrissat, 2014 The
512 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*
513 42:490-495.

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

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

520 MacGregor, E.A., S. Janecek, and B. Svensson, 2001 Relationship of sequence and structure
521 to specificity in the α -amylase family of enzymes. *Biochimica et Biophysica Acta*
522 1546:1-20.

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

525 Marlétaz, F., K.T.C.A. Peljnenburg, T. Goto, N. Satoh, and D.S. Rokhsar, 2019 A New
526 spiralian phylogeny places the enigmatic arrow worms among Gnathiferans. *Current*
527 *Biology* 29 (2):312-318.

528 Philippe, H., R. Derelle, P. Lopez, K. Pick, C. Borchiellini *et al.*, 2009 Phylogenomics revives
529 traditional views on deep animal relationships. *Current Biology* 19 (8):706-712.

530 Pisani, D., W. Pett, M. Dohrmann, R. Feuda, O. Rota-Stabelli *et al.*, 2015 Genomic data do not
531 support comb jellies as the sister group to all other animals. *Proc Natl Acad Sci U S A*
532 112:15402-15407.

533 Plazzi, F., A. Ceregato, M. Taviani, and M. Passamonti, 2011 A molecular phylogeny of
534 bivalve mollusks: Ancient radiations and divergences as revealed by mitochondrial
535 genes. *PLoS one* 6 (11):e27147

536 Pujadas, G., and J. Palau, 2001 Evolution of α -amylase: architectural features and key residues
537 in the stabilization of the $(\beta/\alpha)_8$ scaffold. *Molecular Biology and Evolution* 18:38-54.

538 Ramos-Silva, P., J. Kaandorp, L. Huisman, B. Marie, I. Zanella-Cléon *et al.*, 2013 The skeletal
539 proteome of the coral *Acropora millepora*: the evolution of calcification by co-option
540 and domain shuffling. *Molecular Biology and Evolution* 30:2099-2112.

541 Rödelsperger, C., and R.J. Sommer, 2011 Computational archaeology of the *Pristionchus*
542 *pacificus* genome reveals evidence of horizontal gene transfers from insects. *BMC*
543 *Evolutionary Biology* 11:239.

544 Rodríguez, E., M.S. Barbeitos, M.R. Brugler, L.M. Crowley, A. Grajales *et al.*, 2014 Hidden
545 among Sea Anemones: The First Comprehensive Phylogenetic Reconstruction of the
546 Order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) Reveals a Novel Group of
547 Hexacorals. *PLoS one* 9 (5):e96998.

548 Simion, P., H. Philippe, D. Baurain, M. Jager, D.J. Richter *et al.*, 2017 A large and consistent
549 phylogeny dataset supports sponges as the sister group to all other animals. *Current*
550 *Biology* 27:958-967.

551 Srivastava, M., E. Begovic, J. Chapman, N.H. Putnam, U. Hellsten *et al.*, 2008 The *Trichoplax*
552 genome and the nature of placozoans. *Nature* 454:955-960.

553 Srivastava, M., O. Simakov, J. Chapman, B. Fahey, M.E.A. Gauthier *et al.*, 2010 The
554 *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature*
555 466:720-726.

556 Stam, M.R., E.G.J. Danchin, C. Rancurel, P.M. Coutinho, and B. Henrissat, 2006 Dividing the
557 large glycoside hydrolase family 13 into subfamilies: towards improved functional
558 annotations of α -amylase-related proteins. *Protein Engineering, Design & Selection* 19
559 (12):555-562.

560 Sullivan, J.C., A.M. Reitzel, and J.R. Finnerty, 2006 A high percentage of introns in human
561 genes were present early in animal evolution: evidence from the basal metazoan
562 *Nematostella vectensis*. *Genome Informatics* 17 (1):219-229.

563 Svensson, B., 1994 Protein engineering in the α -amylase family: catalytic mechanism, substrate
564 specificity, and stability. *Plant Molecular Biology* 25:141-157.

565 Tajima, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis.
566 *Genetics* 135:599-607.

567 Uribe, J., Y. Kano, J. Templado, and R. Zardoya, 2016 Mitogenomics of Vetigastropoda:
568 insights into the evolution of pallial symmetry. *Zoologica Scripta* 45:145-159.

569 van der Kaaij, R.M., Š. Janeček, M.J.E.C. van der Maarel, and L. Dijkhuizen, 2007
570 Phylogenetic and biochemical characterization of a novel cluster of intracellular fungal
571 α -amylase enzymes. *Microbiology* 153:4003-4015.

572 Whelan, N.V., K.M. Kocot, T.P. Moroz, K. Mukherjee, P. Williams *et al.*, 2017 Ctenophore
573 relationships and their placement as the sister group to all other animals. *Nature*
574 *Ecology and Evolution* 1 (11):1737-1746.

575 Whelan, S., and N. Goldman, 2001 A general empirical model of protein evolution derived
576 from multiple protein families using a maximum-likelihood approach. *Molecular*
577 *Biology and Evolution* 18 (5):691-699.

578 Wybouv, N., Y. Pauchet, D.G. Heckel, and T. Van Leuwen, 2016 Horizontal gene transfer
579 contributes to the evolution of arthropod herbivory. *Genome Biology and Evolution* 8
580 (6):3594-3613.

581 Zapata, F., F.E. Goetz, S.A. Smith, M. Howison, S. Siebert *et al.*, 2015 Phylogenomic Analyses
582 Support Traditional Relationships within Cnidaria. *PLoS one* 10:e0139068.

583

584 **Legends of figures :**

585

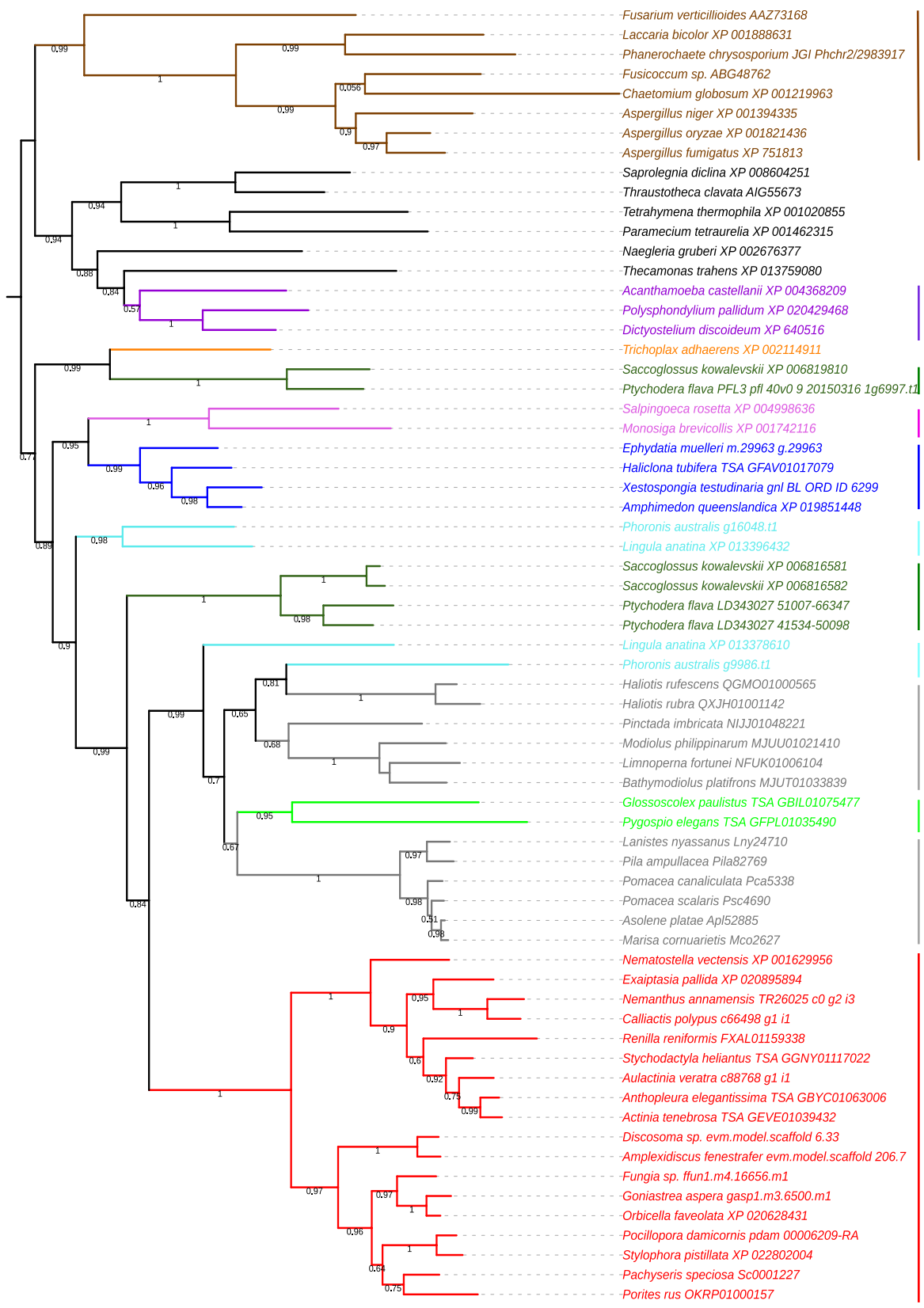
586 **Figure 1:** ML tree of GH13_1 protein sequences of metazoan and non-metazoan species. The
587 tree was rooted by placing fungi and unicellular organisms, except choanoflagellates, as
588 outgroups. The numbers at the nodes are the aLRT supports. Dark green: hemichordates; light
589 blue: brachiozoans;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

592

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

598

599 **Figure 3:** Two scenarii of HGT/gene losses of the GH13_1 genes. HGT or gene loss events were
600 plotted on one of the proposed phylogenies of Bilateria, adapted from Plazzi *et al.* (2011); Kocot
601 (2015); Kocot *et al.* (2017); Luo *et al.* (2015); Luo *et al.* (2018); Uribe *et al.* (2016). Fractions
602 after the lineage names are the number of species showing GH13_1 sequences over the total
603 number of species investigated. A: HGT hypothesis. Black diamonds represent the HGT events,
604 crosses indicate subsequent GH13_1 loss events. B: Gene loss hypothesis. Crosses indicate
605 GH13_1 loss events. Dashed crosses indicate lineages for which only a fraction of the available
606 reliable genome or transcriptome data were found to contain a GH13_1 sequence. Divergence
607 times are from Kumar *et al.* (2017).



Fungi

Amoebozoa

Placozoa

Hemichordata

Choanoflagellata

Porifera

Brachiozoa

Hemichordata

Brachiozoa

Mollusca

Annelida

Mollusca

Cnidaria

- 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
- 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
- Trichoplax adhaerens* XP 002114911
- Saccoglossus kowalevskii* XP 006819810
- Ptychodera flava* PFL3 pfl 40v0 9 20150316 1g6997.t1
- Salpingoeca rosetta* XP 004998636
- Monosiga brevicollis* XP 001742116
- Ephydatia muelleri* m.29963 g.29963
- Haliclona tubifera* TSA GF4V01017079
- Xestospongia testudinaria* gnl BL ORD ID 6299
- Amphimedon queenslandica* XP 019851448
- Phoronis australis* g16048.t1
- Lingula anatina* XP 013396432
- Saccoglossus kowalevskii* XP 006816581
- Saccoglossus kowalevskii* XP 006816582
- Ptychodera flava* LD343027 51007-66347
- Ptychodera flava* LD343027 41534-50098
- Lingula anatina* XP 013378610
- Phoronis australis* g9986.t1
- Haliotis rufescens* QGMO01000565
- Haliotis rubra* QXJH01001142
- Pinctada imbricata* NIJJ01048221
- Mociolus philippinarum* MJJU01021410
- Limnoperna fortunei* NFUK01006104
- Bathymodiolus platifrons* MJJUT01033839
- Glossoscolex paulistus* TSA GBIL01075477
- Pygospio elegans* TSA GFPL01035490
- Lanistes nyassanus* Lny24710
- Pila ampullacea* Pila82769
- Pomacea canaliculata* Pca5338
- Pomacea scalaris* Psc4690
- Asolene platae* Apl52885
- Marisa cornuarietis* Mco2627
- Nematostella vectensis* XP 001629956
- Exaiptasia pallida* XP 020895894
- Nemanthus annamensis* TR26025 c0 g2 i3
- Calliactis polyopus* c66498 g1 i1
- Renilla reniformis* FXAL01159338
- Styrodactyla 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

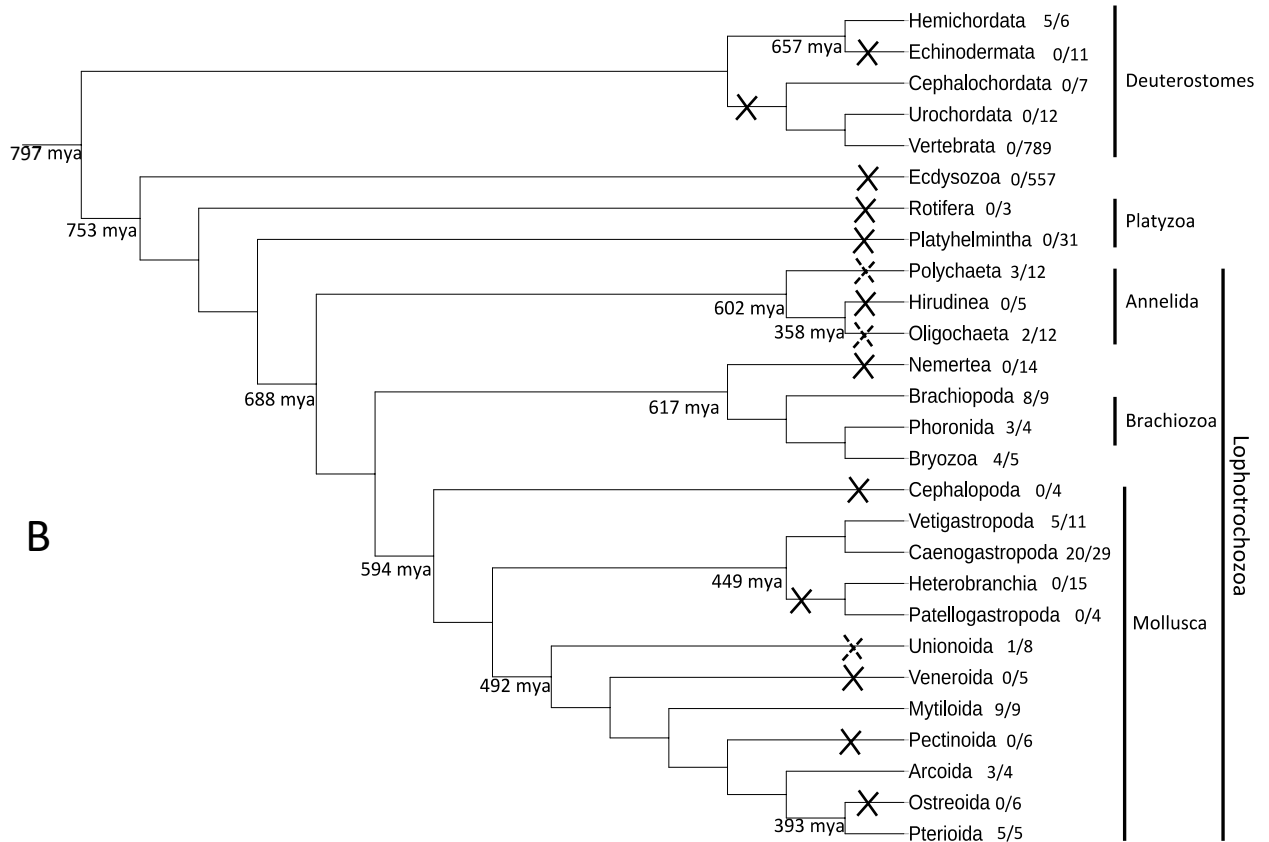
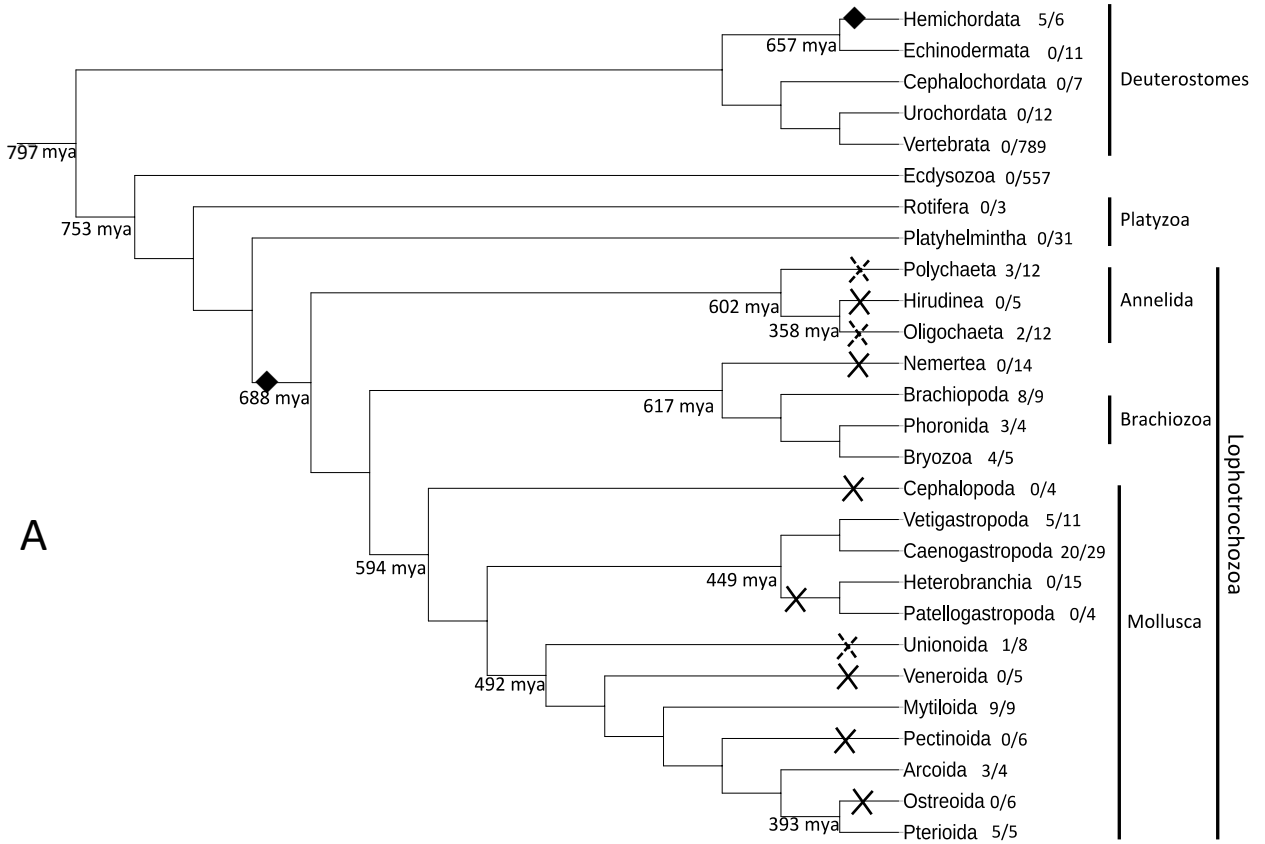


Table 1: GH13_1-like sequences found after BLAST searches in online databases (not comprehensive for unicellars, without the Fungi). * : sequences which have not been characterized as protein-coding, in sequenced genomes with long contigs ; (1) : from short DNA sequences (except Sequence reads archive) ; ** : reported as GH13_1 in CAZy. Most of the SRA data are from transcriptome studies ; see Tables S1 and S2.

Phylum	Species	Database	Accession
NON BILATERIAN METAZOA			
Porifera Demospongiae Heteroscleromorpha	<i>Amphimedon queenslandica</i>	GenBank proteins	XP_019851448
Porifera Demospongiae Heteroscleromorpha	<i>Ephydatia muelleri</i> (1)	Compagen.org	m.29963 g.29963
Porifera Demospongiae Heteroscleromorpha	<i>Haliclona tubifera</i>	GenBank TSA	GFAV01017079
Porifera Demospongiae Heteroscleromorpha	<i>Spongilla lacustris</i>	GenBank SRA	SRX470277
Porifera Demospongiae Heteroscleromorpha	<i>Xestospongia testudinaria</i> (1)	Reefgenomics.org	gnl BL_ORD_ID 6299
Cnidaria Hexacorallia Actiniaria	<i>Actinia tenebrosa</i>	GenBank TSA	GEVE01039432
Cnidaria Hexacorallia Actiniaria	<i>Anthopleura elegantissima</i>	GenBank TSA	GBYC01063006
Cnidaria Hexacorallia Actiniaria	<i>Anthopleura buddemeieri</i> (1)	Reefgenomics.org	c117986_g2_i1
Cnidaria Hexacorallia Actiniaria	<i>Aulactinia veratra</i> (1)	Reefgenomics.org	c88768_g1_i1
Cnidaria Hexacorallia Actiniaria	<i>Calliactis polypus</i> (1)	Reefgenomics.org	c66498_g1_i1
Cnidaria Hexacorallia Actiniaria	<i>Exaiptasia pallida</i>	GenBank proteins	XP_020895894
Cnidaria Hexacorallia Actiniaria	<i>Nematostella vectensis</i>	GenBank proteins	XP_001629956
Cnidaria Hexacorallia Actiniaria	<i>Stychodactyla heliantus</i>	GenBank TSA	GGNY01117022
Cnidaria Hexacorallia Actiniaria	<i>Telmatactis</i> sp.	Reefgenomics.org	C36117_g1_i2
Cnidaria Hexacorallia Corallimorpharia	<i>Amplexidiscus fenestrafer</i> *	Reefgenomics.org	evm.model.scaffold_206.7
Cnidaria Hexacorallia Corallimorpharia	<i>Discosoma</i> sp.*	Reefgenomics.org	evm.model.scaffold_6.33
Cnidaria Hexacorallia Scleratinia	<i>Acropora digitifera</i>	GenBank proteins	XP_015760547 partial
Cnidaria Hexacorallia Scleratinia	<i>Acropora millepora</i>	GenBank proteins	XP_029201467
Cnidaria Hexacorallia Scleratinia	<i>Acropora tenuis</i> *	Reefgenomics.org	aten_0.1.m1.10359.m1
Cnidaria Hexacorallia Scleratinia	<i>Fungia</i> sp.*	Reefgenomics.org	ffun1.m4.16656.m1

Cnidaria Hexacorallia Scleratinia	<i>Goniastrea aspera</i> *	Reefgenomics.org	gasp1.m3.6500.m1
Cnidaria Hexacorallia Scleratinia	<i>Nemanthus annamensis (1)</i>	Reefgenomics.org	TR26025 c0_g2_i3
Cnidaria Hexacorallia Scleratinia	<i>Orbicella faveolata</i>	GenBank proteins	XP_020628431
Cnidaria Hexacorallia Scleratinia	<i>Pachyseris speciosa</i> *	Reefgenomics.org	Sc0001227 74283-80000
Cnidaria Hexacorallia Scleratinia	<i>Pocillopora damicornis</i>	GenBank genomes	XP_027058081
Cnidaria Hexacorallia Scleratinia	<i>Porites lutea</i> *	Reefgenomics.org	plut2.m8.18618.m1
Cnidaria Hexacorallia Scleratinia	<i>Porites rus</i>	GenBank genomes	OKRP01000157
Cnidaria Hexacorallia Scleratinia	<i>Stylophora pistillata</i>	GenBank proteins	XP_022802004
Cnidaria Octocorallia Pennatulacea	<i>Renilla koellikeri</i>	GenBank SRA	SRX4364609
Cnidaria Octocorallia Pennatulacea	<i>Renilla muelleri</i>	GenBank SRA	SRX4717871
Cnidaria Octocorallia Pennatulacea	<i>Renilla reniformis</i> *	GenBank genomes	FXAL01159338
Placozoa	<i>Trichoplax adhaerens</i>	GenBank proteins	XP_002114911
BILATERIA			
Brachiopoda Linguliformea	<i>Glottidia pyramidata</i>	GenBank SRA	SRX731468
Brachiopoda Linguliformea	<i>Lingula anatina</i>	GenBank proteins	XP_013396432
Brachiopoda Linguliformea	<i>Lingula anatina</i>	GenBank proteins	XP_013378610
Brachiopoda Craniiformea	<i>Novocrania anomala</i>	GenBank SRA	SRX731472
Brachiopoda Rhynchonelliformea	<i>Kraussina rubra</i>	GenBank SRA	SRX112037
Brachiopoda Rhynchonelliformea	<i>Macandrevia cranium</i>	GenBank SRA	SRX731471
Brachiopoda Rhynchonelliformea	<i>Hemithiris psittacea</i>	GenBank SRA	SRX731469
Brachiopoda Rhynchonelliformea	<i>Terebratalia transversa</i>	GenBank SRA	SRX1307070
Brachiopoda Phoroniformea or Phoronida	<i>Phoronis australis</i>	marinegenomics	g9986.t1
Brachiopoda Phoroniformea or Phoronida	<i>Phoronis australis</i>	marinegenomics	g16048.t1
Brachiopoda Phoroniformea or Phoronida	<i>Phoronopsis harmeri</i>	GenBank SRA	SRX1121914
Bryozoa Flustrina	<i>Bugula neritina</i>	GenBank SRA	SRX2112329
Bryozoa Flustrina	<i>Bugulina stolonifera</i>	GenBank SRA	SRX6428326
Bryozoa Ctenostomatida	<i>Flustellidra corniculata</i>	GenBank SRA	SRX6428327
Bryozoa Cheilostomatida	<i>Membranipora membranacea</i>	GenBank SRA	SRX1121923
Hemichordata Enteropneusta	<i>Ptychodera flava</i>	Marinegenomics	pfl_40v0_9_20150316_1g2314.t1
Hemichordata Enteropneusta	<i>Ptychodera flava</i>	GenBank WGS	LD343027 41534-50098
Hemichordata Enteropneusta	<i>Ptychodera flava</i>	GenBank WGS	LD343027 51007-66347

Hemichordata Enteropneusta	<i>Ptychodera flava</i>	Marinegenomics	pfl_40v0_9_20150316_1g6997.t1
		GenBank WGS	BCFJ01022326 32811-41459
Hemichordata Enteropneusta	<i>Saccoglossus kowalevskii</i>	GenBank proteins	XP_006816582
Hemichordata Enteropneusta	<i>Saccoglossus kowalevskii</i>	GenBank proteins	XP_006816581
Hemichordata Enteropneusta	<i>Saccoglossus kowalevskii</i>	GenBank proteins	XP_006819810
Hemichordata Enteropneusta	<i>Schizocardium californicum</i>	GenBank SRA	SRX1436000
Hemichordata Enteropneusta	<i>Torquaratoridae antarctica</i>	GenBank SRA	SRX798197
Hemichordata Pterobranchia	<i>Rhabdopleura sp.</i>	GenBank SRA	SRX879690
Mollusca Gastropoda Caenogastropoda	<i>Asolene platae</i>	AmpuBase	Apl52885
Mollusca Gastropoda Caenogastropoda	<i>Batillaria attramentaria</i>	GenBank SRA	SRX2957288
Mollusca Gastropoda Caenogastropoda	<i>Charonia tritonis</i>	GenBank SRA	SRX2753455
Mollusca Gastropoda Caenogastropoda	<i>Conus tribblei (1)</i>	GenBank WGS	LFLW010536118
Mollusca Gastropoda Caenogastropoda	<i>Crepidula novicella</i>	GenBank TSA	GELE01086894
Mollusca Gastropoda Caenogastropoda	<i>Glaussolax didyma</i>	GenBank SRA	SRX5277776
Mollusca Gastropoda Caenogastropoda	<i>Hemifusus tuba</i>	GenBank SRA	ERX3138276
Mollusca Gastropoda Caenogastropoda	<i>Lanistes nyassanus</i>	AmpuBase	Lny24710
Mollusca Gastropoda Caenogastropoda	<i>Marisa cornuarietes</i>	AmpuBase	Mco2627
Mollusca Gastropoda Caenogastropoda	<i>Melanoides tuberculata</i>	GenBank SRA	SRX5832309
Mollusca Gastropoda Caenogastropoda	<i>Neverita didyma</i>	GenBank TSA	GHHQ01002371
Mollusca Gastropoda Caenogastropoda	<i>Nucella lapillus</i>	GenBank SRA	SRX4378318
Mollusca Gastropoda Caenogastropoda	<i>Oncomelania hupensis</i>	GenBank SRA	SRX2739536
Mollusca Gastropoda Caenogastropoda	<i>Pila ampullacea</i>	AmpuBase	Pila82769
Mollusca Gastropoda Caenogastropoda	<i>Pomacea canaliculata</i>	GenBank proteins	XP_025109323 (incomplete)
		AmpuBase	Pca5338
Mollusca Gastropoda Caenogastropoda	<i>Pomacea diffusa</i>	AmpuBase	Pdi16479 (partial)
Mollusca Gastropoda Caenogastropoda	<i>Pomacea maculata</i>	AmpuBase	Pma33988 (partial)
Mollusca Gastropoda Caenogastropoda	<i>Pomacea scalaris</i>	AmpuBase	Psc4690
Mollusca Gastropoda Caenogastropoda	<i>Rapana venosa</i>	GenBank TSA	GDIA01047641
Mollusca Gastropoda Caenogastropoda	<i>Semisulcospira coreana</i>	GenBank TSA	GGNX01073707
Mollusca Gastropoda Vetigastropoda	<i>Haliotis laevigata</i>	GenBank TSA	GFTT01038064
Mollusca Gastropoda Vetigastropoda	<i>Haliotis rubra*</i>	GenBank WGS	QXJH01001142
Mollusca Gastropoda Vetigastropoda	<i>Haliotis rufescens*</i>	GenBank WGS	QGMO01000565

Mollusca Gastropoda Vetigastropoda	<i>Tegula atra</i>	GenBank SRA	SRX958768
Mollusca Bivalvia Mytiloidea	<i>Bathymodiolus platifrons</i> *	GenBank Assembly	MJUT01033839
Mollusca Bivalvia Mytiloidea	<i>Limnoperna fortunei (1)</i>	GenBank Assembly	NFUK01006104
Mollusca Bivalvia Mytiloidea	<i>Lithophaga lithophaga</i>	GenBank SRA	SRX1940727
	<i>Modiolus philippinarum</i> *	GenBank Assembly	MJUU01021410
Mollusca Bivalvia Mytiloidea			
Mollusca Bivalvia Mytiloidea	<i>Mytilus galloprovincialis (1)</i>	GenBank Assembly	APJB011511270
Mollusca Bivalvia Mytiloidea	<i>Mytilus galloprovincialis</i>	GenBank TSA	GHIK01025031
Mollusca Bivalvia Mytiloidea	<i>Perna canaliculus</i>	GenBank TSA	GGLA01150624
Mollusca Bivalvia Mytiloidea	<i>Septifer virgatus</i>	GenBank TSA	GFKS01035611
Mollusca Bivalvia Mytiloidea	<i>Perumytilus purpuratus</i>	GenBank SRA	SRX2210805
Mollusca Bivalvia Mytiloidea	<i>Xenostrobus securis</i>	GenBank SRA	SRX4058936
Mollusca Bivalvia Pterioidea	<i>Malleus candeanus</i>	GenBank SRA	SRX1688295
Mollusca Bivalvia Pterioidea	<i>Pinctada martensi</i> *	GenBank Assembly	CM008066
Mollusca Bivalvia Pterioidea	<i>Pinctada fucata</i>	Marinegenomics	pfu_aug1.0_4142.1_01638
Mollusca Bivalvia Pterioidea	<i>Pteria penguin</i>	GeneBank TSA	GEMO01011007
Mollusca Bivalvia Arcoidea	<i>Anadara trapeza</i>	GenBank SRA	SRX323049
Mollusca Bivalvia Arcoidea	<i>Scapharca broughtoni</i>	GenBank TSA	GEXI01046152
Mollusca Bivalvia Arcoidea	<i>Tegillarca granosa</i>	GenBank SRA	SRX1334524
Mollusca Bivalvia Unionoidea	<i>Cristaria plicata</i>	GenBank SRA	SRX1153631
Annelida Oligochaeta	<i>Drawida calebi</i>	GenBank SRA	SRX6596293
Annelida Oligochaeta	<i>Glossoscolex paulistus</i>	GenBank TSA	GBIL01075477
Annelida Polychaeta	<i>Hydroides elegans</i> *	GenBank Assembly	LQRL01141559
			LQRL01153670
			LQRL01157410
Annelida Polychaeta	<i>Pygospio elegans</i>	GenBank TSA	GFPL01035490
Annelida Polychaeta	<i>Spirobranchus lamarcki</i>	GenBank TSA	GGGS01192599

UNICELLULAR EUKARYOTES

Amoebozoa Mycetozoa	<i>Cavendaria fasciculata</i>	GenBank proteins	XP_004351949
Amoebozoa Mycetozoa	<i>Dictyostellium discoideum</i>	GenBank proteins	XP_640516**
Amoebozoa Mycetozoa	<i>Polysphondylium pallidum</i>	GenBank proteins	XP_020429468
Amoebozoa Discosea	<i>Acanthamoeba castellanii</i>	GenBank proteins	XP_004368209
Choanoflagellida Salpingoecidae	<i>Monosiga brevicollis</i>	GenBank proteins	XP_001742116
Choanoflagellida Salpingoecidae	<i>Salpingoeca rosetta</i>	GenBank proteins	XP_004998636
Ciliata	<i>Ichthyophthirius multifiliis</i>	GenBank proteins	XP_004027176
Ciliata	<i>Euplotes focardii</i>	GenBank proteins	AGU13046**
Ciliata	<i>Moneuplotes crassus</i>	GenBank proteins	AGU13047**
Ciliata	<i>Paramecium tetraurelia</i>	GenBank proteins	XP_001462315
Ciliata	<i>Stentor coeruleus</i>	GenBank proteins	OMJ70617
Ciliata	<i>Stylonychia lemnae</i>	GenBank proteins	CDW84776
Ciliata	<i>Tetrahymena thermophila</i>	GenBank proteins	XP_001020855**
Heterolobosea	<i>Naegleria gruberi</i>	GenBank proteins	XP_002676377
Apusozoa	<i>Thecamonas trahens</i>	GenBank proteins	XP_013759080
Oomycetes	<i>Achlya hypogyna</i>	GenBank proteins	AIG56379**
Oomycetes	<i>Saprolegnia diclina</i>	GenBank proteins	XP_008604251
Oomycetes	<i>Thraustotheca clavata</i>	GenBank proteins	AIG55673**