1	Horizontal transfer and gene loss shaped the evolution of alpha-amylases in
2	bilaterians
3	
5	
4	
5	
6	Andrea Desiderato (1, 2), Marcos Barbeitos (1), Clément Gilbert (3), Jean-Luc Da Lage (3)
7	
8	(1) Graduate Program in Zoology, Zoology Department, Federal University of Paraná, CP
9	19020, Curitiba, Paraná 81531-980, Brazil
10	(2) Department of Functional Ecology, Alfred Wegener Institute & Helmholtz Centre for
11	Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany
12	
13	(3) Évolution, Génomes, Comportement, Écologie. CNRS, IRD, Université Paris-Sud.
14	Université Paris-Saclay. F-91198 Gif-sur-Yvette, France
15	
16	
17	
18	
19	
20	Abstract
21	The subfamily GH13 1 of alpha-amylases is typical of Fungi, but it also includes some
22	unicellular eukaryotes (e.g. Amoebozoa, choanoflagellates) and non-bilaterian Metazoa.
23	Conversely, since a previous study in 2007, all Bilateria were considered to harbor only alpha-
24	amylases supposedly inherited by horizontal transfer from a proteobacterium and classified in
25	the subfamilies GH13_15 and 24, which were therefore commonly called bilaterian alpha-
26	amylases. The taxonomic scope of Eukaryota genomes in databases has been greatly increased
27	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.

35

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

#### 39 Introduction

40 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 41 42 energy to the organism through digestion. They belong to glycosyl hydrolases, a very large 43 group of enzymes which have been classified in a number of families according to their 44 structures, sequences, catalytic activities and catalytic mechanisms (HENRISSAT AND DAVIES 45 1997). Most alpha-amylases are members of the glycoside hydrolase family 13 (GH13), which includes enzymes that can either break down or synthetize  $\alpha$  -1,4-,  $\alpha$  -1,6- and, less commonly,  $\alpha$ 46 47 -1,2- and  $\alpha$  -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 48 49 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-50 51 amylases, while the former was first described in Fungi for which it represents the canonical 52 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 53 54 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, 55 besides the unikont choanoflagellates and amoebozoans, and also excavates and ciliates. They 56 dubbed "Dictyo-type" this alpha-amylase, referring to the slime mold Dictyostelium discoideum 57 (Amoebozoa Mycetozoa). The authors proposed that this amylase, ancestral to the Unikont 58 clade, is shared among non-bilaterian metazoans (e.g. sponges, sea anemones and corals, and 59 60 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. 61

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 nonbilaterian donor(s) which would have to be identified.

69

#### 71 Materials and methods

72 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 73 a query to perform BLASTP and TBLASTN searches on various online databases available in 74 75 Genbank (nr, proteins, genomes, assembly, SRA, TSA, WGS). compagen.org, 76 marinegenomics.oist.jp, reefgenomics.org, marimba.obs-vlfr.fr, vectorbase.org, AmpuBase 77 (https://www.comp.hkbu.edu.hk/~db/AmpuBase/index.php) (IP et al. 2018), between October 78 2018 and April 2019. Fungi were not searched further in this study because they are known to 79 have a GH13 1 member as the usual alpha-amylase. To increase the chances to retrieve potential 80 cnidarian or ctenophoran sequences, the starlet sea anemone Nematostella vectensis amylase 81 (XP 001629956) was also used to query those databases. After the discovery of GH13 1-like 82 sequences in Bilateria, the sequence XP 013396432 of the brachiopod Lingula anatina was also 83 used specifically for additional search in Bilateria. Non-animal eukaryote species were investigated using the Dictyostelium discoideum sequence XP 640516 as query. The BLAST 84 hits were considered further when expectation values (e-values) were better (lower) than  $10^{-100}$ 85 for BLASTP or 10<sup>-75</sup> for TBLASTN, except for constitutively small hits such as SRA (sequence 86 read archives). These were only considered when several highly significant hits covered most of 87 the query sequence. When SRA hits had too many gaps, we did not attempt to assemble longer 88 sequences and thus we did not use such sequences in alignments or phylogenies. Finally, we kept 89 90 only sequences which were inside long contigs, or full-size or near full-size transcripts. We also 91 checked once again the absence of animal-type alpha-amylase (GH13 15 or 24) outside the 92 Bilateria using the sequence of the bivalve Corbicula fluminea (AAO17927) as a BLASTP 93 query. The CAZy database (cazy.org (LOMBARD et al. 2014)), which is devoted to glycosyl 94 hydrolases and related enzymes was used to check assignment of some of the sequences found to the GH13 1 subfamily. 95

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.

100 Alignments were performed using MUSCLE (EDGAR 2004), as implemented in 101 Geneious (Biomatters Ltd.). A maximum likelihood (ML) tree was built using PhyML's 102 (GUINDON AND GASCUEL 2003) current implementation at the phylogeny.fr portal (DEREEPER *et* 103 *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 104 105 stretch. Gaps were removed from the alignments and data were analyzed under WAG (WHELAN 106 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 107 108 sites were estimated from the data. The robustness of the nodes was estimated using an 109 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 110 111 as the ingroup.

- 112
- 113

## 114 **Results**

115 The new sequences retrieved from the databases are listed in Table 1. A general protein 116 alignment of the sequences found in this study along with already known GH13\_1 sequences is 117 shown in Fig. S1.

118

# 119 GH13\_1 sequences retrieved from unicellular taxa

120 We confirmed the presence of GH13 1 in dictyostelids, in ciliates and also in 121 oomycetes, some representatives of which (but not all) are indicated in Table 1. In two 122 oomycetes, Saprolegnia diclina and Achlya hypogyna, the GH13 1-like sequences were the C-123 terminal half of longer sequences, the N-terminal half of which was similar to unclassified GH13 124 sequences found in e.g. Acanthamoeba histolytica (GenBank accession BAN39582), according 125 to the CAZy database. In our general phylogenetic tree (Fig. 1), these sequences were used as 126 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 127 128 Salpingoeca rosetta. A partial sequence was also returned from incomplete genome data from 129 Monosiga ovata (at Compagen, not shown).

130

## 131 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.

139 In sponges (Porifera), data were less abundant. No alpha-amylase sequence was found 140 in Sycon ciliatum (Calcarea) and Oscarella carmela (Homoscleromorpha). All the sequences we 141 retrieved belonged to Demospongiae. Similarly, we found no amylase sequence at all in the 142 phylum Ctenophora (Mnemiopsis leidyi, Pleurobrachia pileus), the phylogenetic position of 143 which is controversial: it has been recovered as the most basal metazoan (WHELAN et al. 2017), 144 as Cnidaria's sister group e.g. (PHILIPPE et al. 2009; SIMION et al. 2017), re-establishing 145 Coelenterata, and also as the earliest branch in the Eumetazoa (animals with a digestive cavity 146 and/or extra cellular digestion) e.g. (PISANI et al. 2015).

147

## 148 GH13\_1 sequences retrieved from bilaterian animals

149 The surprising finding of this study, on which we will focus our attention, is the 150 consistent, albeit sparse, occurrence of GH13 1 alpha-amylase sequences in several bilaterian phyla: hemichordates, which are deuterostomes, brachiopods and phoronids (Brachiozoa) and 151 152 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 153 species, the two copies are located on different contigs. The paralog sequences are rather 154 155 divergent, given their positions in the tree (Fig 1) and each paralog groups the two species 156 together. This indicates that not only duplication, but also the divergence between paralogs is 157 ancestral to these species, dating back at least to basal Cambrian, according to the TimeTree 158 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). 159 160 We must be cautious with transcriptome data alone, since BLAST hits pertaining to transcripts 161 from contaminating symbionts or parasites may be found (BORNER AND BURMESTER 2017). 162 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), 163 but see (MARLÉTAZ et al. 2019) returned no GH13 1 hits, but these animals are still poorly 164

bioRxiv preprint doi: https://doi.org/10.1101/670273; this version posted June 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

represented in sequence databases (only one whole genome in Genbank for Nemertea as of April2019).

167 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 168 169 (XP 006816581 and XP 006816582 in S. kowalevskii, and their counterparts in P. flava) as 170 shown by the topology of the gene tree (Fig. 1). This could suggest independent gene duplication 171 in each species. However, we observed that the two duplicates were arranged in tandem in both 172 species, which would rather suggest concerted evolution of two shared copies. In P. flava, this 173 genome region is erroneously annotated as a single gene at the OIST Marine Genomics database. 174 The three copies were therefore probably already present before the split of the two lineages, 175 some 435 mya (KUMAR et al. 2017). The third paralog is very divergent from the two other 176 copies, so its divergence from the ancestral copy probably occurred before the species split, as 177 well. Another hemichordate species, Schizocardium californicum, harbors a GH13 1 gene, as 178 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, 179 180 personal communication). All the species mentioned above belong to the subphylum 181 Enteropneusta. No data from other hemichordate subphyla were available to us.

182 In molluscs, we found BLAST hits with significant e-values in a few gastropod species 183 from two clades only, the Vetigastropoda (e.g. the abalone Haliotis sp.) and the Caenogastropoda 184 (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 185 186 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 187 188 GH13 1 sequence was found in the gastropod Colubraria reticulata, but in a very short, 189 intronless contig (GenBank accession number CVMW01047267), that was barely longer than 190 the gene itself. Therefore, we disregarded this hit. We also found GH13 1 sequences in a few 191 bivalves, belonging to Mytiloidea (e.g. the mussel Mytilus galloprovincialis), Pterioidea (e.g. the 192 pearl oyster Pinctada imbricata) and Arcoidea (e.g. Scapharca broughtoni). Reciprocal BLAST 193 in GenBank nr using these molluscan high-scoring segment pairs (HSPs) always returned 194 Lingula anatina as the best hit. Although several genomes or transcriptomes have been 195 sequenced from other bivalve and gastropod clades, we retrieved no GH13 1-like sequences 196 from the databases except in the aforementioned clades. We found no such sequence in 197 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.

202 In annelids, we found occurrences of GH13 1 genes in a few species, the genomes of which are 203 still not fully assembled, namely the "polychaetes" Hydroides elegans and Pygospio elegans but 204 not in the well annotated genome of *Capitella teleta*. We also recovered HSPs from the clitellate 205 Glossoscolex paulistus but not from Amynthas corticis or Eisenia fetida. Therefore, in molluscs 206 as well as in annelids, the presence of GH13 1 genes is scattered across lineages. We found that 207 the mollusc GH13 1-like sequences were much shorter, either truncated at the C-terminal, or this 208 region was so divergent from the query sequence (L. anatina) that it was impossible to identify, 209 assemble and align it with our data set (Fig. S1). In addition, we found that the annelid 210 Hydroides elegans had an internal deletion, which precluded its inclusion in the phylogenetic 211 analysis. This suggests that those sequences may not have alpha-amylase activity.

212

## 213 Gene tree analysis: position of bilaterian sequences

214 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 215 GH13 1 alpha-amylase copy ancestral to Unikonts. In the first case, the bilaterians GH13 1 216 217 sequences are unlikely to cluster together and the gene tree topology will likely display one or 218 more nodes that are inconsistent with the bilaterian phylogeny. In the second case, the bilaterian 219 sequences are expected to recover a bilaterian clade and to have a cnidarian clade as its sister 220 group (LAUMER et al. 2018). The actual tree topology (Fig. 1) is not that straightforward when it 221 comes to the bilaterian relationships, although we may rule out any proximity of bilaterians 222 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.

237 Interestingly, the two remaining hemichordate sequences are the earliest diverging lineage of the 238 Metazoa + Choanoflagellata cluster, since they are branched with the placozoan Trichoplax 239 adhaerens sequence, this relationship being strongly supported whatever the tree reconstruction 240 method employed (not shown - Fig. 1). In order to check for the possibility of a long branch 241 attraction (LBA), which would artificially cluster hemichordate and placozoan sequences, we 242 performed Tajima's relative rate tests (TAJIMA 1993) using MEGA7 (KUMAR et al. 2016). The 243 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 244 choanoflagellates. Unexpectedly, the  $\chi^2$  tests returned non-significant values in two tests and 245 significant values in three tests (Table S2). Therefore, with our data, LBA cannot be entirely 246 247 ruled out in this particular case.

248

## 249 Analysis of intron positions

250 Intron positions may be valuable markers when reconstituting gene histories. We 251 identified 56 intron positions from the subset of species of the general tree for which we could 252 find data (Fig. 2). Only one intron position is widely shared among these GH13 1 gene 253 sequences. It is the first position reported in the alignment, and it lies just upstream to the first 254 conserved part of the alignment. The main observation is the numerous conserved positions 255 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 256 257 choanoflagellate Monosiga brevicollis. In contrast, the Cnidaria have few introns, with positions 258 different from the sponge and the bilaterians, except for position 1. The other species under 259 examination, i.e. protists and fungi, have essentially specific intron positions. This is a further 260 argument to state that the occurrence of GH13 1 alpha-amylases in some bilaterians is a story 261 that is internal to metazoans.

262

263

#### 264 **Discussion**

265 We have shown here that a limited number of bilaterian animals, all being aquatic 266 species, namely hemichordates, brachiopods and phoronids, and some sparse molluscs and 267 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 268 269 the GH13 1 sequences we found are not due to contaminating DNA. First, two species with 270 whole genome sequenced and assembled were found to harbor such genes in each phylum 271 Hemichordata and Brachiozoa, and the mollusc Pomacea canaliculata also has a well annotated 272 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 273 274 databases, which added some support to the presence of these amylase genes. Although 275 transcriptome and rough genomic data should be handled with care, this lends support to our 276 observations. Moreover, reciprocal BLAST from the transcriptome hits always returned a 277 bilaterian (L. anatina or S. kowalevskii) best hit, not fungal, protist or other non-bilaterian 278 GH13 1 sequence. Second, the bilaterian sequences retrieved from assembled genomes were 279 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 280 281 is placed at the distal end of a contig, with only two other genes on the contig (Table S1), one of 282 which has a placozoan best hit. But its P. flava counterpart is well inside a very gene-rich contig. 283 Therefore, these seemingly non-bilaterian genes are well in bilaterian genomic contexts.

284 The evolutionary scenario proposed by Da Lage et al (DA LAGE et al. 2007) suggested that the 285 GH13 1 alpha-amylase gene ancestral to Unikonts (Amoebozoa and Opisthokonts, i.e. Fungi 286 and Metazoa/Choanoflagellata) was totally absent from Bilateria, due to complete replacement 287 by a new alpha-amylase, originating from a bacterium through horizontal gene transfer (HGT). 288 The new data unveils a more complicated story. There are two explanations which are mutually 289 exclusive. The first explanation is that several HGTs occurred from non-bilaterian to 290 hemichordate and Lophotrochozoa ancestors. The second explanation is that the ancestral gene 291 was not lost in all bilaterian lineages, but remained (given the current data) in hemichordates, 292 brachiopods and phoronids, and in scattered lineages across Mollusca and Annelida.

293 The hypothesis of HGT requires several such events between metazoans. It implies that 294 HGTs obviously happened after the split of the two main branches of bilaterians, protostomes 295 and deuterostomes, otherwise the transferred copies should have been lost in most phyla, like in 296 the alternative hypothesis. More precisely, in the case of Lophotrochozoa, this would have 297 occurred before the diversification of this clade and after its divergence from the Platyzoa, some 298 700 mya(KUMAR et al. 2017); in the case of hemichordates, after diverging from their common 299 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 300 301 infer at least two HGTs, early in the evolution of each phylum, with subsequent losses in 302 Lophotrochozoa (Fig. 3). The donor species, given the sequence clustering in the trees, could be 303 related to cnidarians. However, we have underlined that the intron-exon structures of the 304 bilaterian sequences were most similar to the one of the sponge, and that the cnidarian GH13 1 305 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 306 307 ancestral structures exemplified by the sponge and the placozoan. Indeed, if the structure shared 308 by the sponge, the placozoan and the bilaterians reflects the ancestral state, cnidarians must have 309 undergone a drastic rearrangement of the intron-exon structure of this gene. This would be in 310 line with the long internal branch leading to this clade in the trees (Fig. 1), which suggests 311 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.

318 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 319 320 (HEJNOL AND MARTINDALE 2008; FLORES AND LIVINGSTONE 2017). In this respect, the 321 GH13 15/24 gene that was acquired from a bacterium is a type of horizontal transfer akin to 322 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 323 324 (HUSNIK AND MCCUTCHEON 2018) (see also (DA LAGE et al. 2013)). In contrast, while numerous 325 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 326 327 events have been inferred that involve a metazoan donor and a metazoan receiver (GRAHAM et 328 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 329 330 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 331 comprehensive taxon sampling, turn out to be rather due to recurrent gene losses (HUSNIK AND 332 MCCUTCHEON 2018). The conservation across phyla of the intron-exon structure, probably 333 334 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 335 336 T. adhaerens (SRIVASTAVA et al. 2008).

337

338 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 1 339 subfamily. Indeed, some of them have been assigned to this subfamily in the reference database 340 341 CAZy.org (see Table 1), and if we add sequences from the closest subfamilies, namely GH13 2 342 or GH13 19 (STAM et al. 2006) in the alignment and in the phylogenetic tree, the putative 343 GH13 1 and the ascertained GH13 1 remain well clustered together (not shown). It is possible 344 that modifications of a few amino acid positions could bring a change in the substrate or catalytic 345 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 346 found in scleractinians and molluscs (MARIN et al. 2007; RAMOS-SILVA et al. 2013). Given the 347 calcium binding activity of alpha-amylases (BOEL et al. 1990; GROSSMAN AND JAMES 1993; 348 349 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 350 351 biomineralization process. In the analyses performed by those authors, no amylase was found in 352 the shell matrix, but this does not exclude the possibility of its presence in the pathway. The fact 353 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 354 355 function, would add credence to another function. This conjecture requires further investigation. 356 On the other hand, the full-size GH13 1 sequences only present in a few bilaterians could have 357 remained true alpha-amylases with the classical function, but this would make even more 358 enigmatic why they have been conserved, either by descent or by horizontal transfer.

bioRxiv preprint doi: https://doi.org/10.1101/670273; this version posted June 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

- 359
- 360

361 Acknowledgments: We want to thank Pedro E. Vieira for introducing AD to molecular 362 analyses.\_We are grateful to Didier Casane and Emmanuelle Renard for fruitful advise and 363 discussion. We thank Héctor Escrivà for sharing sequence data. This work was funded by the 364 Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (process no. 365 141565/2017-9) to AD and regular funding of the CNRS to JLDL and CG. CG is also supported 366 by a grant from Agence Nationale de la Recherche (ANR-15-CE32-0011-01 TransVir).

367

# 368 **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 Calciumbinding 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.
- 379 Cordaux, R., and C. Gilbert, 2017 Evolutionary significance of Wolbachia-to-animal horizontal
   380 gene transfer: Female sex determination and the f element in the isopod *Armadillidium* 381 *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.
- 385 Da Lage, J.-L., E. G. J. Danchin and D. Casane, 2007 Where do animal α-amylases come from?
   386 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: 114.
- 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.
- 418 Kocot, K. M., 2015 On 20 years of Lophotrochozoa. Organisms Diversity and Evolution 16:
  419 329-343.
- 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: 256282.

- 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: 18701874.
- 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.
- 431 Letunic, I., and P. Bork, 2016 Interactive tree of life(iTOL) v3: an online tool for the display
  432 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:
  435 490-495.
- 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.
- 445 Marin, F., G. Luquet, B. Marie and D. Medakovic, 2007 Molluscan shell proteins: primary
  446 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.
- 452 Pisani, D., W. Pett, M. Dohrmann, R. Feuda, O. Rota-Stabelli *et al.*, 2015 Genomic data do not
  453 support comb jellies as the sister group to all other animals. Proc Natl Acad Sci U S A
  454 112: 15402-15407.

- Plazzi, F., A. Ceregato, M. Taviani and M. Passamonti, 2011 A molecular phylogeny of bivalve
  mollusks: Ancient radiations and divergences as revealed by mitochondrial genes. PLoS
  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.
- Ramos-Silva, P., J. Kaandorp, L. Huisman, B. Marie, I. Zanella-Cléon *et al.*, 2013 The skeletal
  proteome of the coral *Acropora millepora*: the evolution of calcification by co-option
  and domain shuffling. Molecular Biology and Evolution 30: 2099-2112.
- Rodríguez, E., M. S. Barbeitos, M. R. Brugler, L. M. Crowley, A. Grajales *et al.*, 2014 Hidden
  among Sea Anemones: The First Comprehensive Phylogenetic Reconstruction of the
  Order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) Reveals a Novel Group of
  Hexacorals. PLoS One 9: e96998.
- Simion, P., H. Philippe, D. Baurain, M. Jager, D. J. Richter *et al.*, 2017 A large and consistent
  phylogeny dataset supports sponges as the sister group to all other animals. Current
  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.
- Sullivan, J. C., A. M. Reitzel and J. R. Finnerty, 2006 A high percentage of introns in human
  genes were present early in animal evolution: evidence from the basal metazoan *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.

497

## 498 Legends of figures :

499

**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.

506

**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.

512

**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

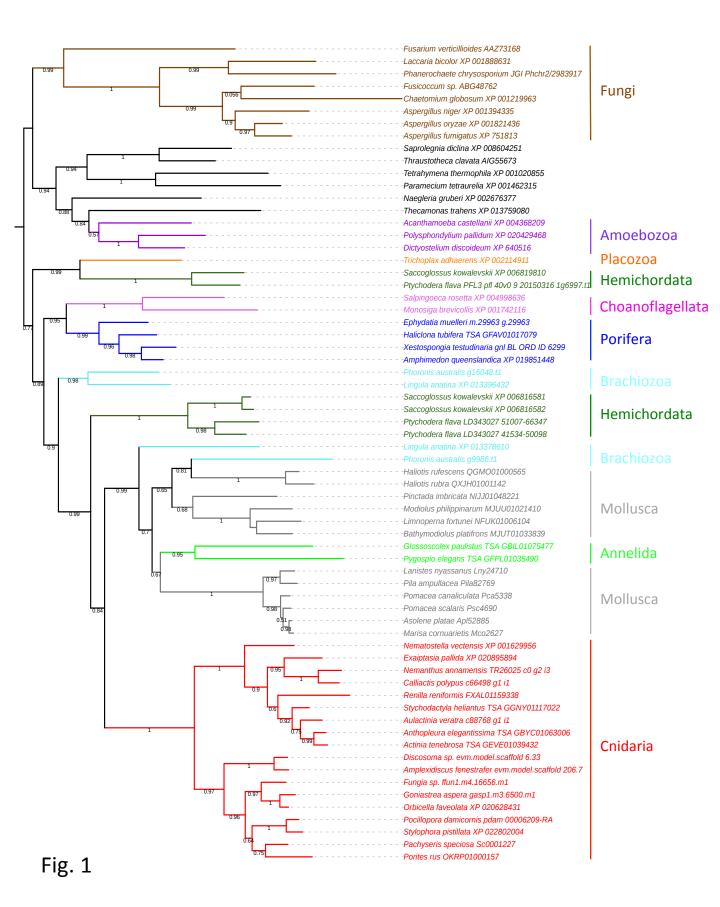
- 524 Legends of Supplementary figures
- 525

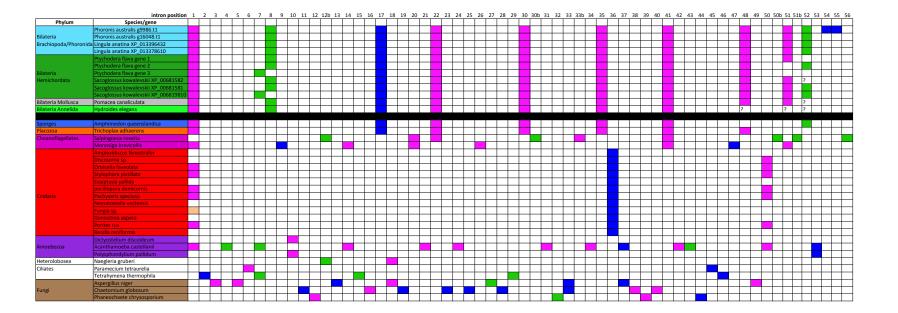
526

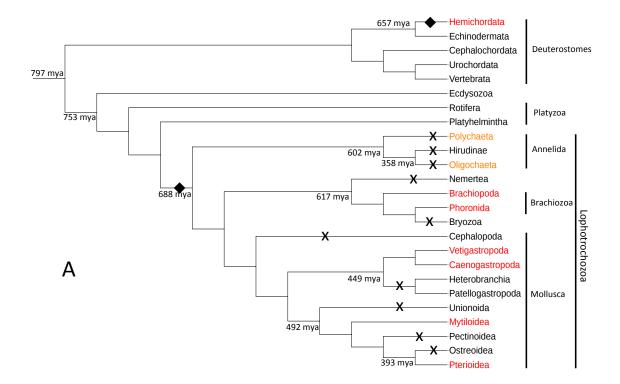
- 527 Figure S1: General protein alignment of the GH13\_1 sequences used for the gene tree.
- 528 Background colors indicate the conservation level of amino acids : red : 100% identity; orange :
- 529 80 to 100% identity ; yellow : 60 to 80% identity ; white : below 60% identity.

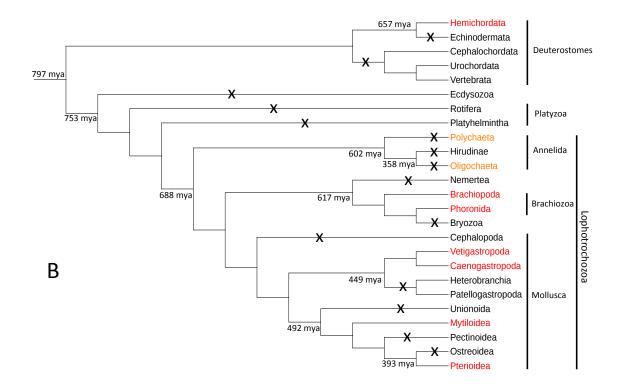
530

531









**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	L.		
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	$c88768 \ \overline{g1} \ \overline{i1}$
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 Cnidaria Octocorallia Pennatulacea Placozoa

### BILATERIA

Brachiopoda Linguliformea Brachiopoda Linguliformea Brachiopoda Linguliformea Brachiopoda Phoroniformea or Phoronida Brachiopoda Phoroniformea or Phoronida Brachiopoda Phoroniformea or Phoronida Brachiopoda Craniiformea Brachiopoda Rhynchonelliformea Brachiopoda Rhynchonelliformea Brachiopoda Rhynchonelliformea Hemichordata Enteropneusta

Hemichordata Enteropneusta Hemichordata Enteropneusta

Hemichordata Enteropneusta Hemichordata Enteropneusta Hemichordata Enteropneusta Hemichordata Enteropneusta Mollusca Gastropoda Caenogastropoda Mollusca Gastropoda Caenogastropoda Mollusca Gastropoda Vetigastropoda Mollusca Gastropoda Vetigastropoda Mollusca Gastropoda Vetigastropoda Mollusca Gastropoda Caenogastropoda Mollusca Gastropoda Caenogastropoda Mollusca Gastropoda Caenogastropoda Stylophora pistillata *Renilla reniformis\** Trichoplax adhaerens

GenBank proteins GenBank genomes GenBank proteins

XP 022802004 FXAL01159338 XP 002114911

Glottidia pyramidata	GenBank SRA	SRX731468 (transcriptome)
Lingula anatina	GenBank proteins	XP 013396432
Lingula anatina	GenBank proteins	XP_013378610
Phoronis australis	marinegenomics	g9986.t1
Phoronis australis	marinegenomics	g16048.t1
Phoronopsis harmeri	GenBank SRA	SRX1121914 (transcriptome)
Novocrania anomala	GenBank SRA	SRX731472 (transcriptome)
Kraussina rubra	GenBank SRA	SRX112037 (transcriptome)
Macandrevia cranium	GenBank SRA	SRX731471 (transcriptome)
Terebratalia transversa	GenBank SRA	SRX1307070 (transcriptome)
Ptychodera flava	Marinegenomics	pfl 40v0 9 20150316 1g2314.t1
	GenBank WGS	LD343027 41534-50098
Ptychodera flava	GenBank WGS	LD343027 51007-66347
Ptychodera flava	Marinegenomics	pfl_40v0_9_20150316_1g6997.t1
	GenBank WGS	BCFJ01022326 32811-41459
Saccoglossus kowalevskii	GenBank proteins	XP_006816582
Saccoglossus kowalevskii	GenBank proteins	XP_006816581
Saccoglossus kowalevskii	GenBank proteins	XP_006819810
Schizocardium californicum	GenBank SRA	SRX1436000
Asolene platae	AmpuBase	Apl52885
Conus tribblei (1)	GenBank WGS	LFLW010536118
Haliotis laevigata	GenBank TSA	GFTT01038064
Haliotis rubra*	GenBank WGS	QXJH01001142
Haliotis rufescens*	GenBank WGS	QGMO01000565
Crepidula novicella	GenBank TSA	GELE01086894
Lanistes nyassanus	AmpuBase	Lny24710
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
Monusca Gastropoda Caenogastropoda	1 omacea canaiicaiaia	AmpuBase	Pca5338
Mollusca Gastropoda Caenogastropoda	Rapana venosa	GenBank TSA	GDIA01047641
Mollusca Gastropoda Caenogastropoda	Semisulcospira coreana	GenBank TSA GenBank TSA	GGNX01073707
Mollusca Bivalvia Mytiloida	1	GenBank Assemby	NFUK01006104
5	Limnoperna fortunei (1)	5	MJUT01033839
Mollusca Bivalvia Mytiloida	Bathymodiolus platifrons*	GenBank Assembly	
Mollusca Bivalvia Mytiloida	Modiolus philippinarum*	GenBank Assembly	MJUU01021410
Mollusca Bivalvia Mytiloida	<i>Mytilus galloprovincialis (1)</i>	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	GEM001011007
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_013396431	XP_023931706	XP_013396389	XP_023931696	XP_023931697	XP_013406295
XP_013396432	No animal hit (°)	Vertebrata e-57	Bivalvia e-11	Echinodermata e-	Hemichordata 0.0	No animal hit
				48		
Lingula anatina	XP_013378627	XP_013378628	XP_013378631	XP_013378654	XP_013378655	XP_013378611
XP_013378610	Annelida e-53	Bivalvia e-118	Gastropoda e-59	Annelida e-51	Cnidaria e-63	Cephalochordata e-33
Saccoglossus	XP 006816579	XP 006816580	XP 006816576	(1)	(1)	(1)
kowalevskii	No hit (°)	Hexapoda e-73	Cephalochordata			
XP_006816581			e-60			
XP_006816582*						
Saccoglossus			XP_006819811	XP_006819812		
kowalevskii			No hit (° )	Placozoa e-47		
XP_006819810						
Phoronis australis			g16047.t1	g16049.t1	g16050.t1	g16051.t1
g16048.t1			Annelida e-71	No hit	Bivalvia 0.0	Brachiopoda e-36
Phoronis australis	g9983.t1	g9984.t1	g9985.t1	g9987.t1	g9988.t1	g9989.t1
G9986.t1	Brachiopoda e-62	Bivalvia 0.0	Brachiopoda e-95	Gastropoda e-77	Echinodermata e-	Brachiopoda 0.0
					76	
Ptychodera flava		1g2312.t1	1g2313.t1	(1)	(1)	(1)
LD343027 41534-		No hit	Hemichordata e-			
66347*			60			
Ptychodera flava	1g6994	1g6995	1g6996	1g6998	1g6999	1g7000
1g6997	No hit	Hemichordata 0.0	Hemichordata e-	No hit	Hemichordata e-	Echinodermata e-
			79		79	62

Pomacea	XP_025110260	XP_025109347	XP_025111249	XP_025109322	XP_025110021	XP_025110160
canaliculata	Mollusca e-33	Annelida e-64	Mollusca e-91	Mollusca e-80	Mollusca e-145	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	$\chi^2$	р
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

er na for tunei NFUK01006104	TOREATEAEHEVTCTDU-RS-WCGTE			110 120 130 DMV ANHMG YNHGCCWNVE C SOKOL VN- FT	140 150 GENPESEAGHYH	160 170 180 190		10 RLAK
rubra QXIH01001142	NITOTEKG-DNPNTTCSDL-RS-YCGGSF	KG EEOLDY OGLGANALWISPILENTVN	GYHGYWAKDIYNINPLEVTEDDLKFLISECH KREIWYMY GYHGYWASNINKVNPEFGTEAELKSLITECH KRDIWYM	DVVANHMGRPSGCIDRSCPTEELTN-FT	EFAPEDROEHYH	-0LCY	GSSNOSEVEOC	RLGI
rufescens QGM 001000565	LLTDRFSQKG-DNPNTTCSDL-RS-YCGGTF	KG EENLDY QGLGANA WISPTVENTVN	GYHGYWAKNIYKVNPEEGTEEELKSLIAECHKRDIWVMF	DVVANHMGRTSGCSDRSCPAKKLTN-FT	EFVPFDQQEHYH	-EL <mark>C</mark> Y <mark>1</mark> Q	GS – – – – – – S <mark>NQ S EME</mark> E <mark>C</mark> I	RLGI
imbricata NIJJ01048221			GYHGYWAKDIYRVNPEEGTEEDLVDLIRACHGRDIWVML					
odiolus platifrons MJUT01033839	LLTDRFATENGKEMPCNDL-RS-YCGGTF	RGIAEKLDYIS <mark>GLG</mark> A <mark>NAIWISPIV</mark> VNTPK	G <mark>YHGY</mark> WAKDIYSINPHEGSENDLKFLITECHKRDIWVMV	DVVANHMGHTHGCFWNIDCNP-QIVN-ET	G <mark>F</mark> N <mark>PFN</mark> ETD <mark>HYH</mark>	-DL <mark>C</mark> E <mark>1</mark> K	NWKNQTEMEIC	RLAK
is philippinarum MJÜU01021410	L FTDRFASDDKETSCNDL-RS-YCGGTF	RG AQKLDY SGLGANA W SPVVENTPK	GFHGYWAKDIYSINPHEGTKDDLKYLISECHKRDIWVMV	DVVANHMGYNDGCCWN NCTHKQLVN-ET	GEIPENETSHYH	-PLCE	NEKNQTEMEVC	RLAK
a scalar is Psc4690	LLTDRFSPSG-AAPSQPCTDL-RN-YCGGTF	RG AQHLDY QGLGANA WISPFVENNPN		DVVANHMGVAPGDPEDLS	GEVPENESWHYH	-TSCQIQ		RLEN
oullacea Pila82769	LLTDRFSPTG-PPPDQPCADL-RN-YCGGTF LLTDRFSPSG-AAPSQPCTDL-RN-YCGGTF		····NYHGYAAKNIYAVNPMEGTRQELLNLIRACHERDIWVM ····NYHGYAAKNIYAVNPMEGTKQDULDUTRDCHQRDIWVM			-APCQIQ -TRCQIQ		
cornuarietis M co2627 nyassanus Lny24710						-TP <mark>C</mark> Q <b>I</b> Q		
platae A pl52885	LLTDRFSPSG-AAPSOPCTDL-RN-YCGGTF		NYHGYAAKNIYAVNPMEGTKQDLLDLTRDCHQRDTWVM					
a canaliculata Pca5338	LITORESPSG-AAPSOPCTDL-RN-VCGGTE	RGVACHLDY LOGI GANALWISPEVENNPT	DYHGYAAKNIYAVNPMEGTKODLLDLTRDCHORDTWVM		GEVPENESWHYH	-THCI		RIEN
olex paulistus TSA GBIL01075477	LITORESOSD-VDPSEPCRDL-RS-WONGTW	KGLIGHLDY ION GVDALWISP VENTEG	GYHGYWAKNLNRLNDHEGSKEDLLALVEECHRRDTWVM	DVVPNHMGYPMNCSYME CPEEYRNN-ER	FETPENET HYH	- FYCF1	TWPTTNOTFLEVC	RLAY
egans TSA GFPL01035490	LTDRESKPG-DGTDAACVDL-SD-WCGGNE	NG VORLOY LEELGATATWISPETLNTPG	GYHGYWAKNLYEVDPNYGTEAELLNLIEKCHERDTWVM	DVV PNHMGYPHDCKWPNPPTCDETKLVN-ET	GLTPEDOAEHYH	- E ICNIE		RLAN
atina XP 013396432	LITOREHNPS AGGSCGNL-RD-YCGGTE	OGVIKKLDY OGLGVNALWISPVVENTPG	GYHGYWAKN SN NPKEGSPODLKDLVSACHARG WWW	DVVGNHMGYPPGCPGCSMOOETD-ES	SLWPENRAEHYH	-NYCDIK	NWGDAHEVEYC	RLAE
atina XP 013378610	ILTORESPHG-PTPSTPCTAL-RS-YCGCKE	KGLOORLDY KGLGANALWISPIVLNTPG	GEHGYWAKNTYETNPEEGTKQDLKDLTKACHDKDTWVMV	DVV PNHMGYNDGCFWNNPCTPTKLNN-ET	GEVPEDKPEYYH	-QYCE1M	DWNNOTENE FC	RLAH
ustralis g9986.t1	ILTDRFSHTD-TDPEQQCSDL-RS-FCGGTF	KG QKRLDY TGLGANA W SPVVENTPG	GYHGYWAKNIFKINPHEGTEDDLKELIAECHKRDVWVMV	DVVANHMGYPGSCNWTDTCTDAQLQD-YH	DFVPFNETE <mark>HYH</mark>	- SLCHIR	NWNNQTEVELC	RLAN
ustralis g16048.t1	TDRFARSD GNTDNCSDL - RN - YCGGAF	KG VNNLDY EGLGVNA WESPVVEQTDG	GYHGYWAKDLTKINSHEGTADDLKNLVAECNKRDIWVMV	DVVANHMGYPPGCPGCSASANVN-FT	GFNPFNKAEYFH	-NYCIIN	DWSNQGEMENC	RLAD
flava LD 343027 41534-50098	LTDRFAHTE-QSPDTECAEL-YE-YCGGTL	KGIQNNLDY AGLGANAIWISPFVENTEL	GEHGYWPKNINKVNSNEGTEQDLKDLIAACHERDIWVMA	DLVLNHMGYQHGCHWSGCPLEKTDN-FT	GFVPFDKTEHYH	- P F C D 1	DWSKEEPEN <mark>WQ</mark> GC	.W <mark>LA</mark> Y
flava LD 343027 51007-66347	ILTDRFAHAD-QSPDTECKEL-YE-YCGGTF	KGIQNNLDYISGLGVNAIWISPITENTKL	GEHGYWPKNIYEVNPHEGTEQDLKDLIAACHERDIWVMA	DLVLNHMGYQEGCHWTGCPVEKTDN-FT	G <mark>F</mark> T <mark>PFD</mark> DAK <mark>YYH</mark>	- P Y <mark>C</mark> D <mark>V</mark>	DWQNEREENVQGC	WLAY
flava PFL3_pfl_40v0_9_20150316_1g6997.t1			DYHGY SAIDMYTI SWHWGTEK DFNDLVAAAHARDTWIMV					
us kowalevskii XP_006816582	LIDRFAHTD-KSPDEDCQEL-YE-YCGGTF	KGTENNEDN NGLGANATWESPFAENTEL	GEHGYWSKNIYETNPHEGSKODLKDLIAACHSRDWWW	DLVLNHMGYPDGCHWSGCPIDNANN-ET	EFVPENDEADYH	- I YCD	DWSNEQQENVEMC	REAY
is kowalevskii XP_006816581	LIDREAHID-KSPDEECQEL-YE-YCGGIE	KGIENNLDY IGLGANATWISHFIENTEL	GEHGYWPKNIYETNPHEGSKQDLKDLIAACQSRDWWWA	DLVLNHMGYPDGCHWSGCPIDNANN-EI	GEVPENDEADYH		DWSNEQUENNE C	RLAY
is kowalevskii XP_006819810	LIDRESAPG-SASWIPCGDL - TE-YCNUNE		DYHGY SAIDFYKVSWHWGT SQDFFDL AAAHARDIW MV					
1uelleri m.29963 g.29963	LLTDRFAQNQDSSYPCGDL-SN-YCGGTF		GYHGYWQKDIYTINPSEGTSDDLKSLVNECHSRNIWVML					ALAN
lon queenslandica XP_019851448 a testudinaria gnl  BL_ORD_ID 6299			GYHGYWQQD ITQ INPNEGTEQDLKDLVNACHSKDIWVML GYHGYWQQNLTQ INPNEGSEQDLKDLVQACHTRDIWVML					
bifer a TSA GFA V01017079			GYHGYWQQD TQ NPHEGTSDDLKNLVDTCHKRD WVML					
adhaerens XP 002114911	VLVDRFAGYN-SNSWSSCKDL-KG-YCGGSF		GYHGYWAKD YK INSHYGTSQELKDLVDAHKQGVWIML					
cus fenestrafer evm.model.scaffold 206.7			GYHGYWAKD I FQ INPHEGTKNDLK SL I KACHDRDVWVM					WLAG
sp. evm.model.scaffold 6.33	TTREAPSG-DLPSOSCOL-RD-YCGGTE	KGIEKHLDY TGLGANALWISPLVVNTDK	GYHGYWAKD I FQINPY FGTKDDLK SLVKACHDKDVWVW	DVVANHMGYPPDTSWHVDGGTTKVLDBED		-PPHPY		WLAG
veolata XP 020628431	LI TOREAPSG-EVPSOKCVGL-SN-YCGGTE	KGLEKHLDY TGLGANALWISPLVI NTDN	GYHGYWAQNIFEINPHEGTKEDLKSLVAACHDRDVWVMV	DIVANEMGYPPNTDPLPDVSMSEN	SEVPENDSKYEH	-PVHPY	KWPEECKDOWKTELY	WAN
la heliantus TSA GGNY01117022	LTDREAPSG-NLPATKCTDM-RG-WCGGTE	KG ONHLDY TGLGANA W SP VLNTEK	GEHGYWAKN FE EPHEGTKEELKALVKACHERDVWVMV	DVVANHMGYPPGTDWTTP WNSTLFDRFYE	NEHSENRSEYEH	- PNHP Y	KWPEECHDVKK OEY	WLAN
pisti ata XP 022802004	LLTDREAPSG-DLPPDKCGDL-SN-YCGGTE	KGIEKHLDY TGLGANA WISPLVONTDK	GYHGYWAQNIFEINPHEGSRDDLKSLVTACHERGWWML	DLVANHMGYPPEASFN	DL PENESSHYH	-PKHPY	KWPOECKDOWKEENY	WLAG
pallida XP_020895894	LTDRFAPSG-DLPTTKCTNL-RG-WCGGTF	KG QXHLDY SGLGANA WISP VLNTNG	GYHGYWAKN FELESHEGSKEDLKNLVKACHERDIWVM	DVVANHMGYPPNTDWTTPWNSTLFNNFPD	HETPENKXEYYH	-PQHPY	KWPEECHDLHK QTY)	'W <mark>LA</mark> N
veratra c88768 g1 i1	ILTDRFAPSG-SLPTTKCTDM-RG-WCGGTF	KGIENHLDYITGLGANAIWISPIVLNTEK	GFHGYWAKDIYEIETHEGTKEDLKSLVRACHERDVWVMV	DVVANHMGYPPGTNWETP WNSTLFDKFYE	NLY <mark>PEN</mark> RSE <mark>YYH</mark>	-PNHPY	KWPEECHDVKKIQEY	'W <mark>LA</mark> N
ebrosa TSA GEVE01039432	LTDRFAPSG-SLPTVKCTDM-RG-WCGGTF	KG QNHLDY TGLGANA WISPIVLNTDK	GEHGYWAKNIYEIESHEGTKEELKSLVRACHERDVWVMV	DVVANHMGYPPGTNWNTPWNSTLFDRFYE	NLY <mark>PFN</mark> RSE <mark>YYH</mark>	-PNHPWI	KWPEECHDLKKLQEY)	'W <mark>LA</mark> N
OKRP01000157			GYHGYWAQN FEINPHEGTEEDLKSLVKACHDRDVWVMA					
damicornis pdam_00006209-RA	LLTDRFAPSG-DLPSRKCDDL-KN-YCGGTF	KGIEKHLDYITGLGVDAIWISPIVQNTDR	GYHGYWAQNIFEINPHEGSRDDLKSLVAACHERGVWVML	DLVANHMGYPPGTGFN	NLIPENESRHYH	-PQHSYI	KWPQECKDQWKTENY	WLAG
speciosa Sc0001227	LLTDRFAPSG-ELPPQKCSDL-KD-YCGGTF	KGIEKHLDY TGLGANA WISPIVQNTDK	GYHGYWAQNIFEINPHEGTKEDLMSLVKACHERDVWVML	DLVANHMGYPPNADWTRKLEN-FD	KFVPENDAKYEH	-PYHSY	KWPEECKDQWK <b>LE</b> NY	WLAN
a vectensis XP_001629956	LLIDRFAQSG-EMPAKKCIDM-RG-WCNGIF	KGIEKHLDY IGLGANAIWISPIVLNIDQ	GEHGYWAKNIYEIEPHEGTKQDLKSLVKACHDRGVWVMV	DVVANHMGYPPGVDWRTPWNSSLLDNFYE	YFYPENKSEYYH	-ANHK Y	KWPEECHNLIK QKY	WLAN
un1.m4.16656.m1	LLIDREAPSG-ELPSQKCVGL-RN-YCGGIE	KG EQHLOY IGEGANATWISH VLNTVN	GYHGYWAQN IFEVNPKEGTREDLKSLVTACHDRDWWWA	DVVANHMGYPPNINWNIDPSPNVSMSFU		- PRHP Y		MLAN
aspera gasp1.m3.6500.m1 a elegantissima TSA-GBYC 01063006						PVHPY		
lypus c66498_g1_i1			GYHGYWAKDLFK TE SHEGTKEDLK SL KACHERDWWW					
annamensis TR26025[c0_g2_i3			GYHGYWAKDIFKIEPHEGTKEDLKSLIQACHERDWWW					
amanensis 1720023100_g2_13			GEHGYWAKS YE EPHEGTKEDLKSLVKACHDRD WWW					
m discoideum XP 640516	LITORESOTV NSSOPCONE-OG-VCGGTE	OGVEAHLDX OGMGEDA W SPVVTNTPG	GYHGYWQQD I YTVNEY EGTENDLLNM KACHERG WVML	DVVANHVG-PVNVDVG	TIVPEDSVEHVHNCTTC	PQYCTD		RISG
lium pallidum XP 020429468	TOREAL TNNATTPOPDI-TT-YCGGTE	VOVENHEDY LOGMO FDATWISP LVTNTPG	GYHGYWQQDTYTTNSNEGTDNDLLNLTGECHKRGTWVML	DVVANHVG-PVDEDVG		P SQCT NVCI ITNLIFQHY I DID I VVVVCYSLK		
ba castellanii XP_004368209	LI TOREAOTS - DGSYAPCNNL - SN-YCGGSE	OG INHLDY IGMG FDALWISPLVTNTPG	GYHGYWSQDLFGINPHEGTADDLRQLISACHARGVWVML	DVVANHMGGNVPMS	OLS-PENOSEHYBRCVAC	P SNCQ1N		KLAG
evicollis XP 001742116	LTDRFANPSOPTEACADL-ST-YCGGTW	AAGVPYLDY RKLGATAVW SPIPDNTDG	GYHGYWOONMSRLNANFGDANGLTOFVEACHKLEKVML	DVV INHVGNODSSDHNDFH	OFYPENSVDHYH	- SYC    0	DETNLOOVOLC	RLSG
rosetta XP 004998636	ILTDRFAQSS-SGSTAPCNNL-QS-YCGGTW	NGTARMAQYIKQLGANALWISPIPENTDN	GYHGYWOKN YNLNPNYGTGDOL TOM NOL OOMD WVML	DVV INHMGNODNGK LNDFS	MEYPENDSSHYH	- SY <mark>C</mark> Q <mark>1</mark> Q		RLAN
n tetraurelia XP 001462315	LLTDRYASTK SNSTSC-NL-GN-YCGGNF	KG EKNLDY EDLGFDA W SP VQNYDG	AYHGYAAKNIYQINEHEGTAQDEKDEVNACHQKDIWVM	DIVANHMGNODLDFS	RNYPENSSSHEH	-DLCE1S	DTDF ATHNLVNERC	RLFM
a ther m ophila XP_001020855			QEHGYAAKNMTEINSHEGGADGLKSLIDACHKRDIWVMI					
ca clavata A IG55673	VLTDRFAGGSRSDL-GS-YCGGTF	N <mark>GL</mark> KN <mark>NL</mark> K <mark>Y  </mark> KD <mark>MG</mark> F <mark>DA  W   SPVT</mark> D <mark>N  </mark> DN	GYHGYWFKNFEKINSHEGTADDLKNLVATAHQMDTWVMV	DVVANHVG-PVGTNYA	QIY <mark>PEN</mark> SSS <mark>HYH</mark>	- S SCD1T	DWKNQNM <mark>ME</mark> NC	RLSS
diclina XP_008604251	LTDRFAGPSSGACKDL-ST-YCGGTF	N <mark>GI</mark> KS <mark>NLQYI</mark> R <mark>GMG</mark> F <mark>DAIWISPVV</mark> DNSPG	G <mark>YHGYWARN</mark> WDKINANEGTPEELKALVDAAHALDMWIMV	DVVANHVA-PVGNDES	SIVPENNPSHYH	- P SCE	DWSNQKMVETC	RLSG
s trahens XP_013759080	VLTDRFGGPV-TDSC-NL-ND-YCGGTF	SGLIDHLDY EGMGFDALWESPVVTNTPG	GYHGYWAKDFSTINSHEGSAADLKALVDAAHRRDMWVML	DIVLNHVG-PVGSQFS	AITPESHPGDYH	-PQCSVS	QWVCFTEEMLHC	RLAS
ruberi XP_002676377	I TDRFSPSE-SGSGKSCPLL-YT-YCGGTF	QGTISHENNYKSEGFNALWISPIVLNLPG	GYHGYWAQDWFAVNPHEGTQADL INLVKTAHSMDTWVM	DVV FNHVG-PVGFDFS	EIST-PENSSTHYH	-PYCQ	NWSNQTEMEYC	RLAD
fum igatus XP_751813	LLTDRFARTD-GSTTASCDTS-AR-YCGGTW		EGTAYHGYWQQDIYSVNSNYGTADDLKALASALHDRGMYLMV	DVVANHMUYAGAGDSVDYS	VENPENSQISEH	-PLCFI		WEGDNSV
oryzae XP_001821436			YGDAYHGYWQQDIYSLNENYGTADDLKALSSALHERGMYLMV			-PFCFIQ		MEGDNIV
niger XP_001394335			DGE AYHGYWQQK I YDVN SNEGT ADDL K SL SDAL HAR GMYLM					
rertici∥ioides AAZ73168	ALTDRVARSADDNGGSACGNL-GN-YCGGTF VLTDRFARTD-GSTTATCNTQDRV-YCGGTY		·····GYHGYWAKDLYAVNSKIYGTADDLKSLVKSAHDKNMYVMO YGQAYHGYWQQDLYRLNEN FGSADDLKALAAELHDRDMYLMN					
n sp. ABG48762 m globosum XP 001219963	VVTDRFARTD-USTFATCNTQDRV-TCGGTY		ADGS SYHGYWQODLYRLNENEGSADDLKALAAELHDRDMYLM MDGS SYHGYWAKD IWALNAAFGDENDLVELSAELHARDMYLM					MOGS
ete chrysosporium JGI Phchr 2/2983917	LVTDRFATTDGSSPPCDTSNRO-YCGGTW	OGVISKIDY ONMERDAVWISPVVKNI FONTS	YGEAFHGYWTQDPTQLNSHFGSENDLKAL SDALHKRGMYLM			-SECEINAT	DYPSPPYSAGNOTR	WEGDTNU
icolor XP_001888631			YGQAYHGYWAQD T SLNPH FGTADDLKAL SSAL HKRGMYLM					

	230 240 250 260	270 280 290 300 310	320 330 340 350 360 370	380 390 400 410 420 430 440
Lim noper na fortunei NFUK01006104				LŚSH-VSILMNATTYVELGOG <sup>–</sup> TPTTYYGTEÓHENGADDPHNREALW-POYNTST-AMYK
Haliotis rubra QXJH01001142 Haliotis rufescens QGM 001000565		QLSAGVECIGEANYYDRPRYLAAYQGEALDALLNFPMYWA		MTPD-VTLENNALVEVEFS OG - UT NYVYTEQUENEGED BINNESLT = NEDNTT-AMYO MTPD-VTLENNALVEVEFS OG - UT NYVYTEQUENEGED BINNESLT = NEDNTT-AMYO LSR-H-TTLENNALTVULEOGE - UT NYVTEQUENEGED BINNESLT = NVNTS-KIKKEN LSR-H-TTLENNALTVULEOGE - UT NYVTEQUENEGED BINNESLT = NVNTS-KIKKEN LSR-H-TTLENNALTVULEOGE - UT NYVTEQUENEGED BINNESLT = NVNTS-KIKKEN LSR-H-TTLENNALTVULEOGE - UT NYVTEQUENEGED BINNESLT = NVNTS-KIKKEN LSR-H-TTLESNALTVULEOGE - UT NYVTEQUENEGED BINNESLT = NVNTS-KIKKEN LSR-H-TTLESNALTVULEOGE - UT NYVTEQUENEGEN BINNESLT = NVNTS-KIKKEN LSR-H-TTLESNALTVUENEGEG - UT NYVTEQUENEGEN BINNESLT = NVNTS-KIKKEN LSR-H-TLESNALTVUENEGEG - UT NYVTEQUENEGEN BINNESLT = NVNTS-KIKKEN LSR-H-TLESNALTVUENEGEG - UT NYVTEQUENEGEN BINNESLT = NVNTS-KIKKEN LSR-H-TLESNALTVUENEGEG - UT NYVTEQUENEGEN BINNESLT = NVNTS-KIKKEN 
Pinctada imbricata NIII01048221	SNEYWRDSLK SWI KWLG-DEWEFDGYR I DTAVEWDK PFWSEFO	DSSAVRECINGEANNGDNTTYLASYOG-PLDSVLNEPLEHA	IVDVECKNKSMSVEIDSLOEOVTNDVTILGTEVDNHDFPRELS	
Bathymodiolus platifrons MJUT01033839	TNEYNRKTLLNWIRDLT-QKYQIDGYRIDTAVENAKDEWTDE	QIAAGAYCVGEANNGDNMTYVAGYQG-PLNAVLNFPVYWS	LRNVFTENQPMTLEKTTLEREVTLEKDVTVLGNELDNHDFQRELN	LSRH-TTLLMNSTTYVLLGQG-TPITYYGTEQLLHGGNDPENREALW-PNYNTST-KMYK
Modiolus philippinarum MJUU01021410	TNEYWRSTLLKWIRDLT-KEYDIDGYRVDTVTEVLSEFWTEF	QE <mark>A</mark> <mark>AGAY</mark> CLGEANNDDNMT YVAE YQG - PLNAVLNEPVYWA	IRHVENENKSMTILKTTLEQEVRKEKDVTVLGNEIDNHDFPRELY	LSPH-ETVLMNALTEVLLAQVLVPLIYYGTEQLYHGEYDPKNREALW-PNYNTSS-GMYR
Pomacea scalar is Psc4690 Pila ampullacea Pila82769		WASSGVEMLGEANNGDRTCYTGGYOG-PLPSVLNEPLYWA WASSGVEMLGEANNGDRTCYTGGYOG-PLPSVLNEPLYWA		
Marisa cornuarietis M co2627	TEPWVRRTLLNW ADTT-REVGEDGYR IDTVVEVEK PEWSQEV	WASSGVEMLGEANNGDRTCYTGGYQG-PLPSVLNFPLYWA	MRRVENERQPMTELSESLEAQAACESDLTVLGLEADNHDFPRELN	LSYD-TTLUSNALTYVMFGEG-TPLINYGTEGEYHGGNDPYCRESLW-PNYNTNA-EMYKFISTUTH-
Lanistes nyassanus Lny24710	TVPWVRHALLDWIANIT-HEYGFDGYRIDTVVEVEKSEWSQE/	NA <mark>S</mark> <mark>SGVE</mark> MLGEAN <mark>NG</mark> DRTC <mark>Y</mark> TGA <mark>YQ</mark> G - PLP <mark>SVLNEPLYW</mark> A	MRRAENEROPMTELTKSLEAQAACESDLTVLGLEADNHDFPRELN	LSYD-TSLLSNALAYYLFGEG <mark>-IPI I YYGTEQ</mark> EYHGGNDPYC <b>RESLW-P</b> NYNTEA-PMYKFYSKLTR-
A solene platae A pl52885 Pom acea canaliculata Pca5338		WASSGVEWLGEANNGDRICYIGGYQG-PLPSVLNEPLYWA		LSYD-TTL USNAL I YVL F GEG- PTL YYG I E GEWHG GNP YCR ESLW - PNYNTNA-OMYK FI STL TH-
Glossoscolex paulistus TSA GBIL01075477	TVPENROTLYDW RNMT-I ENGEDGYRVDTVRH EMGEWPGYC			
Pygospio elegans TSA GFPL01035490	TVP EVRQLLLDWIRNTT-TTYK FDG YRVDTVRHIELPFWGEY	QAASNNELLGEVWNEDPFYVAAYQE-QMDSLFNFPLHEQ	MRQVFTYRAQNMTFTADGLVQQRAVFRDTQVLGVFVDNHDKPRFLN	FTQD-LAL <mark>LKNAVTESLLTDG-TRFWYYGTEQAY</mark> SGGADECC <mark>RESLW-PNYNT</mark> AS-D <mark>MEKEHQTL</mark> IR- INKD-WTS <mark>LKNALAYTLM</mark> ARW- <b>TPIVYYGTE</b> QAYAGGNDEY <mark>NRESLW-PNYNT</mark> NH-L <mark>MYQFI</mark> AQ <mark>V</mark> NA-
Lingula anatina XP_013396432				
Lingula anatina XP_013378610 Phoronis australis_g9986.t1				WTSD-YTVLKNAL TYVILGEG-IPI I YYGTEQGE SGGHDPNDRESLW-PYYSTQS-DI YKFI AGLTK-
Phoronis australis g16048.t1	ENPYWRTTUKTWYHDUV DTYKEDG IR IDT I PEWPKDEWSEE	SKSAGVEOLGEVDNGDPSYVGPYOG-PLTAVLNYPFFFT	FROVEGNERS MRG TORYGOE SOHEO DLS LGGEVONHONPRELS	GNDD-WTALKNALAEN MAOW-TPT TYYGTEGAEK GGNDPNNRE TLW-PHYDNKH-LMYOFTRTWTS-
Ptychoder a flava LD343027 41534-50098	TNTENRQTLLDWTTDUT-TRYGFDGYRLDFVKGWSMDFWQEVD	DE <mark>A</mark> CKE <b>Y</b> T <mark>IGE</mark> AY <mark>DG</mark> NLD VV <mark>A</mark> S <mark>YQ</mark> G - NLD <mark>SVL</mark> S <mark>FPMYF</mark> V	LEDLEQNNPSTGSMER HEQLELEATAYQDATILGGELDNHDRARELS	KNSD-VTALKNALVYLLMARW-IPFVYYGTELGFKGGNDPQNRESLW-PHVDTEH-ELYKFIAKVNK-
Ptychoder a flava LD 343027 51007-66347		DDACQSYTIGEAADARLDVVSPYQG-PLDSALSEPMYYV		RNSD-VTALKNALVYLLMARW-IPIIIYYGTELGFIGGDENDRESMW-PHVDTNH-ELYKFIAKUNE-
Ptychoder a flava PFL3_pfl_40v0_9_20150316_1g6997.t1 Saccoglossus kowalevskii XP_006816582				
Saccoglossus kowalevskii XP 006816581	TVASWRETLLNWVREFN-AEWGFDGYRLDFVKGVSKDFLAEVD	DEASDSETLGEAYDGRTDYVASYQG-YLDSLLSEPMYET	LEDVEOTEPL-GKLTQ HDOLLREADEYQDTSVLGGELDNHDRKRELA	THE STATE AND A ST
Saccoglossus kowalevskii XP_006819810	ENPYMOTELCTWIDWFI-KTEDFDGLRIDTVSHWPKWEWEHF	YA <mark>A</mark> <mark>ANTY</mark> E <mark>VGE</mark> VYTSDLD KTQS <mark>YT</mark> P - P <mark>MDA</mark> ALNYPL YWA	MKDAF IWGGTMRD ANMYRDEKYRFGDGLHLFGNFADNHDQARFLC	DKDS-WTLLKNYIMLVLTMEG-IPIVYYGTEQGYDCCADPNNRESLY-PNEDTNH-HLYLFIQQRTHQ
Ephydatia mueller i m.29963 g.29963 A mphimedon queenslandica XP 019851448	SNSFNQSTUISMIHNII-SINGFDOINVDIIPENPKSFWYQYS	SGSAGVIISMGEVENGDPEYVSGYQG-PLIAILNYPMYEK		NNPD-RVTLKNALAYVI FAEG-IPI I YYGTEQGED SSDPYCRESLY - PKYNTSS-DLYLFINTLAT-
Xestospongia testudinaria gn  BL ORD ID   6299	SNS FVR STLKNWIKE OV OKYDLDGLRVDTTPEVPKDFWAEYS	SOSASVETIGEVENGDENYCSGYOG-PL SAVL DYPMYYK		QNSD-WNTLQNALAWNIFAES- <b>TRI HYYSTEO GEN D</b> YKDENNRESLW - PHYNTDH - PLYKAIFSIWAK KNSD-WNTLQNALAWNIFAEG-HARG- <b>TRI HYYGTEO GR</b> SBSSODENNRESLW - PHYNTGH - LYYGSIATULK - QNSD-WNTLLANAWTUFAEG-HEI HYYGTEO GRSBSDENNRESLW - PHYNTGH - LYYGSIATULK -
Haliciona tubifer a TSA GFA V01017079	DNS FVR STLKQWI HNT I - STYGFDG I RVDT TPENHPDEWSEYS	SSS AGVYSIGEVFNGDPN YVSK YQG - PLSATLNYPMYYK	LKDAFQSKSSMRETHDGVNANS-VFSDASILGTEIDNHDNPRELS	QNSD-WNTLLNAMAYTIFAEG-IPITYYGTEQGES@PSDPNNRESLW-PHYDTTN-TLYTKISIMAK-
Trichoplax adhaerens XP_002114911	ENNEWKNELKYWYSWIV-KEYNEDGIRIDTIPEIPVWEWSEE	HTSANVETIGEAFNGFVDYLKKYVP-PLDSVLNYPMYYA	FKDVYMYKQSMTRIRDVLNEEYSKFSDVQLLGLELDNHDVPRFLY	FNND-WTLLKNNLAWILLGEG-IPIVYYGTEQAYAGGNDPNNRESLF-PNENQNH-WMYDYIKTIND-
A mplexidiscus fenestrafer evm.model.scaffold_206.7 Discosoma sp. evm.model.scaffold_6.33				INDK-RSGUMINELIEVLMADW-IPVYYGIEGGINGGADYNNKESUW-PNMSIES-PLYKLIQNLAI- INDK-PTCIMB-ITVYWADW-IPVYYGIEGGINGGADYNNKESUW-PNMSIES-PLYKLIQNLAI-
Orbicella faveolata XP 020628431	SHPFWHOTLLKWIODLT-TEFEFDGYRLDTTIOVPKLFWAEF	RKAGGVEWMGEANNGPPPCGSMEYVGQYQE -Y DSVLAYPMYWT	MRWIFQEKSKDFTSESKAVNNSYDIYKDRSVFGGEVDNHDNPRELS	MNPS-QTALKNELALVIMGDW-IPIVYYGTEQGENGGNDPNNRESLW-PFISAEN-PLYKFIETLIK-
Stychodactyla heliantus TSA GGNY01117022	SHPFMEKTLLDWIQFLI-NEWDFDGCRVDTVIQVPKPFWTKE	QE <mark>A</mark> GK <b>VE</b> MLGEANNGPPPCGSLNETAGFQG-PLNSVLDFPLFWT	NRY I EQEKKQ DFLSESKYLKDSAKRYL DRSVIGGEVDNHDHERELH	— FIND WTLKINN AND FILLES FIND WTLKINN AND FILLES FILL
Stylophor a pistillata XP_022802004 Exaiptasia pallida XP_020895894				
Aulactinia veratra c88768 g1 i1	SHPFWEKTLLEWIKYLI-KEYDFDGCRVDTV IOVPKDFWTKF	OKAGKVFMLGEANNGAPPCGTLNFTAGFOG-PLNSVLDFPLFWT	NRY I FGEKKONFTSLSKYLKESSKRFLDRSVLGGF I DNHDHARFLH	KNP 5 - YTALRNNU VY IFMSQC- <b>TRI LYYGTEQGE SGOND</b> RIC <b>RESLW - P</b> YMNPKH - VLYQFTRNU IQ- LNP 5 - YVT <mark>LRNNU VYLFM</mark> SQS- <b>TRI LYYGTEQ</b> G <b>E</b> N <b>G</b> DGDRN <b>NRESLW - P</b> YMN <b>T</b> NH - E <b>LYTFT</b> KR <mark>L</mark> I K -
A ctinia tenebrosa TSA GEVE01039432	SHPYMEKTLIHW OYLV-NEWDEDGCRVDTV OVPKPEWTKEC	NKAGKVEMLGEANLGPPPCCALNETAS EQG-PLNSVLDEPLEWT		
Porites rus OKRP01000157 Pocillopora damicornis pdam 00006209-RA		RKAGGGEMIGEANNGPPPCGSMEYVAGYQG-YIDSVLAYPSYWT	LRK FQEKSKDFNALSVAVKNAYQ YKDRSV GGEVDNHDNPRELN	LNP S-HTALRNAU TYVLMADW- <mark>TRTVYYGTEQCENGOS DRNNRESLW-P</mark> EMSTEN-PLYKFTOTLVH- LNP S-HTS <mark>LKNAU AFLMM</mark> GDW- <b>TRTLYYGTEQ</b> GENGGSDRN <b>NRESLW-P</b> SLS <b>T</b> AN-P <b>LYKFT</b> OTLAK-
Pachyseris speciosa Sc0001227	SHPFWHONLINWIHNET-NOFFFDGYRIDTTIOVPKDFWTFFF	RKAGGA WIGEANNGPPPCGSMETVAETOG-TOSVIATPMIN	LRELEOFKTKDETSESSAVKKSYELYKDRSVFGGEVDNHDNPRELN	
Nem atostella vectensis XP_001629956	SHPFNEKTLLDWIKWLI-TENDFDGCRVDTVIQVPKPEWTKE	QS <mark>A</mark> G <mark>GVE</mark> MLGEANNGPPPCGT I NETAP EQG - PLDSVLDEPMEWT	LRY IF QEKTQNFTSESKALKESSKAFKDRSILGGEVDNHDHERFLH	KNPS-QTSERNNEAFVEMSRW-TPLTYYGTEQGENGGGDPNNRESLW-PFMDRKN-SLYVFTKDLTA-
Fungia sp. ffun1.m4.16656.m1	SLPFWRETLLKWIDDLT-NEEGFDGYRLDTTIOVPKPEWAEF	RKAGGVEMIGEANNGPPPCGSMEYVAQYQR-YVDSVLAYPMYWT	LRW FORKSKNFRSENVALNKSYELYKDRSVLGGEVDNHDNPRELH	
Gonlastriea aspera gasp1.m3.6500.m1 Anthopleura elegantissima TSA GBYC 01063006		RKAGGVHWWGEANNGPPPCGSMEYVGQYQQ-YIDSVLAYPMYW DKAGKVHWIGEANIGPPPCGTINFTAGFOG-PLHSVLDEPLFWT	NRY FOENKONESSISKY KDSAKAEL DRSV GVEDNHDNPKELS	WNPS-QUALRNE AFVINGDW-PLIVITGIEQGENGODANNKESLW-PYFSVEN-PLIVITGIIK-
Calliactis polypus c66498 g1 i1	SHPYWANVLLNWVKNTV-KEWDIDGYRIDTVIQVPKWFWSEY	QKAGGVFMLGEANNGPPPCGTVNFTAGFQG-PLNTVLNFPLFWT	MRYILQEHTQNFFSLSKNLKDSAASFQDRSILGGEIDNHDHERFLY	KNPS-YTSLRNNLVYLFMSQC-TPVLYYGTEQGENGGNDPHCRESLW-PHMNTSS-KLYKFTKQLIQ-
Nem anthus annam ensis TR26025 [c0_g2_i3	SHPFWAKVLLDWVKNTV-EEFDIDGYRIDTVIQVPKWFWSKY	QK <mark>A</mark> GR <b>VE</b> MLGEANNGPPPCGTVNETAD EQG-PLNSVLNEPLEWT	MRY I EQEHTK NFFS SKTLNDTAAAFE DRSILGGEIDNHDHERELY	KNPS-YSALRNNUVYLFMSKC-IPVLYYGTEQGFSGQNDPYCRESLW-PYENTSS-TLYKFIKHVVR-
Renilla reniformis FXA L01159338 Dictyostelium discoideum XP 640516		JKAGKVHMLGEANNGPPPCGTLNLTSSFQG-DLDSVLDFPLYWT	MRY EQENPGELDENALSDLLKRSSKKYK====DRSVIGGEIDNHDHARELH===	INPN-YVSERNN AYLLMSQC-IPVLIYYGIEQGENGASDENRESLW-PYMNISH-VEIFIEKKUVG- COSD-IEIVKNA-MAQG-IPVLIYYGIEQGENGASDENRESLW-PYMNISH-VEIFIEKKAS-PLYOFUNT-
Polysphondylium pallidum XP 020429468	NNTEWREAL VNWIKNVT-EFYGEDG IR IDTVPEWDODFWIEYT	TTAAGVY SVGEVYNGDVEYVASYQP-VIDGLL SYPLFF	LR SVEAQQQSMYQ QSMLQQYNATGLHQQYLGTEIDNODQIRELN	EQSD-IELYKNAUTYVLMAQG-LELIYYGTEQGENGASDENNREPLWTTSENIIAS-PLYQEIQTWYT- EQSD-IELYKNAUTYILTSEG-LEIYYYGSEQAENGGADEDNRETLWTTGFNIDS-TLYQFIRTWVL-
A cantham oeba castellanii XP_004368209	TNPYWKSQLINWIRTLMDYGFDGLRIDTVPELEPGFWVDF	VSAANTYAVDLSCCATYQKQALPGVLSYPLFWT	LRSVEQQKQSMNQLQSTFFSYQSQEADMDLLGTEIDNHDNARELS	GTSD-YKLYQNAITYT MAQG-TPTTYYGTEQGYDGTNDPNNRESLWQTRFSTQN-PLYQLIRTIVS- INSD-SVALANALAWILFAEG-TPTTYYGTEQGESGGNDPANREDLWRSGYQTSN-A <mark>lf</mark> tlLTVLTT-
Monosiga brevicollis XP_001742116 Salpingoeca rosetta XP_004998636	SNTEMDGELROW KSTL-QTEGEDGLEYDIVMENPAAYWASY ENGEWATTEONWULDVL-NDWULDVLERDVLEWDVDLWKAW		LRHAFQEKQSMWGLSTTWKDCKASEQDTKLLGLEVDNHDNPRELN	INSD-SVALANALAWILFAEG-LEITYYGTEOGESGGNDPANREDLWRSGYQTSN-ALFILLTVLTT- IRND-YCALRNALAYILFAEG-LEITYYGTEOGEAGGNDPANREDLWRSNYDTSA-PVFOFIATLTA-
Paramecium tetraurelia XP_001462315	DNSYMANEL NW KWV SEEN DG R DTAMWKGEFWKKFT			
Tetrahymena thermophila XP_001020855	ENSYMSNYLVNW KOLV-QTYNVDG IR IDTVPENHPDEWKQYT	TTSAGVEATGEVEDGDEGYVASYIP-SVGSVLNYPWYE	MRD FVNQKDMWGWRTYYQNWANQHADLSLLTPFVDNHDNPRFLSDQV	FNNGKDRNTR - IQLUKGYTTFTLTS IG-IPILWYGTEQYFSGNTDPYDRE PLW-NSLNTNS-DMYKYYQSTLY-
Thraustotheca clavata A IG55673	DVSFWRSYLKNWVKQLV-QTYGFDGIRIDTVPEWHPDFWAEY	35 <mark>5</mark> <mark>AGVE</mark> Q <b>VGE</b> VFNGDPAYVGPYQK-YLT <mark>SV</mark> FNYPMYFT	ADVEGSGMDMSNLQNRYNAEGSYEKDIDALGSEVDNHDNARELN	KFPEKKAQLTAANVFVLTGRG-LEFVYYGTEQYENGDADPNNREVLW-TKMDTSS-DFYKLFKVVNA- KYPNKQAQLTAATVFSLTTEG-LEFVYYGTEQYYGGGVDPANRESLW-TNLDTTS-DLYKILSVVHR-
Saprolegnia diclina XP_008604251 Thecamonas trahens XP_013759080				
Naegleria gruberi XP_002676377	SNNFWYKTLSNWTTDFVLNTEGFDG RIDTVPELSKTEWSDL	KTN I LAPKYGD I YS I GEVL NGDMNYVAPYAN-I LGATL HYPMYFV	LKSIFSQGGSMYQ RNTINSARSLVKDTSLLGVEIDNHDNPRELN	YESD-YKL VONALTYVET SEE YESD-YKL VONALTYVET SEE YKD-MALAKINAAAF IT SDD YTKD-MALAKINAAAF IT SDD YTKD-MALAKINAAAF IT SDD YTKD-MALAKINAAF IT SDD YTKD-KALKINAAF INDE 
A spergillus fumigatus XP_751813	TNPDWQKIWYNWWNSLV-SNWSIDGLRIDTVKHWQSDEWPGEN	ND <mark>A</mark> <mark>AGVY</mark> C <mark>IGE</mark> VFDGDPA <mark>Y</mark> TCP <mark>YQ</mark> E - VLDG <mark>VLNYPIYY</mark> F	LLKAFQSTSGSMSSEYDMINTVKSQCADSTLLGTEVENHDTPRFAS	YTKD-MALAKNAAAEILESDG- <b>TELLYAQOGOHYSGGADPANREAVI</b> LSGYSTTS-DLYKLIATANA-
Aspergillus oryzae XP_001821436 Aspergillus niger XP_001394335		NKAYQN-VMDGVLNYPLYYP DEMAGVMCVGEVDNGNPALDCPYQN-VMDGVLNYPLYYP		
Fusarium vertici lioides AAZ73168	G S D T WK K V L Y D W K WL Y - S EN S F D G I R I D T V K H W E K P F W P G F C		MNREYOOKGDPSDWVAMHDEISSKEPDPTILGTEIDNHDNPRWLS	
Fusicoccum sp. ABG48762				
Chaetomium globosum XP_001219963	V SEAWR SVWNEWVTEWV - SKY SIDGL RVD SAKHVETSEWSGE		LNAFKSPDGSISEVSGUNTLRGVAQDPSUYGSULENHDIERFPS	FTED-TALVKNCSNRPWKG- <b>TETINCCOECHYSC</b> SGTPONREALWSSGYSETA-ELYTWIAKLNO- QTSD-QSLIMNAMAWPFVQDG-VPIMYYCO <mark>E</mark> CSYACGADPANREALWLSGYQIDK-PLVKHAQALNA-
Phanerochaete chrysosporium JGI Phchr2/2983917 Laccaria bicolor XP 001888631	ENQUE VINEWINT BELONWITE TO MAN VET UT VKHLKOP VIDE ENPSTVSTMNNW OGLVEKSWGIDGIRIDTAKHWRKOF WPAR	SVDAGVETMGEVLANCICYVGPYTK-FLPAVLDYPTWFF AMSYVAPYTO-VLDSVLDYPAVYT	TSAFSSOTGNISAFISSVTASOKSYSTPSVVGSFFNHD0PCOS	ITQD-ISLVKNAMAWPFVQDG-WEINYYQQEGSMALGADPANKEALWLSGYQIDK-PLVKHAQALNA- ITQD-ISLVKNAMTWPFIQDG-IPILYQGQEQGFTGGVVPANQEALWTTGFNTSNKPLVTHVTTLNA-

Lim noperna for tunei NFU K01005104 Halicis rufes (QM-0010026 Halicis rufes cers QM-0010026 Battym odda site is Sec4900 Pla am pulacea Pla82709 Molosis philiparerum MUU01021410 Pernaces calastis res.c4990 Pla am pulacea Pla82709 Molosis philiparerum MUU01021410 Fernaces calastis res.c4990 Pla am pulacea Pla82709 Molosis philiparerum MUU01021410 Sectorea Pla82709 Molosis philiparerum MUU01021410 Pernaces calastis res.c4990 Pla am pulacea Pla82709 Molosis philiparerum MUU01021410 Pernaces calastis res.c4990 Pla am pulacea Pla82709 Molosis philiparerum MUU01021410 Pernaces calastis res.c4990 Pla94 Pla94 Pla94 Pla9470 Asclere plate Apt52855 Pla94 Pla94 Pla9470 Percors autralis gl998611 Am philiper autoris gl998612 Percors autralis gl99861 Am philiper autoris gl99861 Percors autralis gl99861 Percors autralis gl99861 Percors autralis gl99863 Percors autralis gl99863 Percors autoris philiper autoris gl99863 Percors autoris philiper 200398980 Percors autoris philiper autoris gl99863 Percors autoris philiper 200398980 Percors autoris philiper Percors autoris gl

							1.04				-		0000		-
- TVRVDLMAKN-C	R <mark>K</mark>				LVAL	INVGRGVS		SHLA I					-RDG5		GMPOV
-ATRVDQMAKN-C															
-AVHVDFTARN-C	RTV	YADRDLL	LESRG	GAEL	LVAL	TNVGRGHN-	-LOTS	KRLPL	QDAD	-YVSV	U	SEVHLS	-RDGA	ALVVMDN	GIPOT
ATRVDLMAKN-C	ROI	FADHDLL	LFSRG	√-GSE <mark>L</mark>	LVAL	TNVGRGVS-	L R K S	SRLAV	ODGE	-YVSV	F I	SKVRLS	-RSGS		GMPOV
-TARVDLTAKN-C	R <mark>Q</mark> I	FVDRDLL	LFSRG	E – G S E <mark>L</mark>	LVAV	TNVGRGVS-	L R K S	SRLAV	QD <mark>G</mark> D	-YISA	F	SKVRLS	-RGES	CE <mark>MVL</mark> SD	GMPQV
- NLLSTVISTT-C	EEI	YADDDTL	VETKG	NQ	IVVV	INL GTGS S	QQK	SFAVAQ	QDGT	EYEN	H	SSDVVIN	-RNGH		GOPKV
-KLGSAFYDSA-C			AFTRO			SNOGSGST-									GOPKI
-SHUSHEVSGG-C	REM.	AANDDEE	VETRG	K-FNOF	VVI	TNRGNTSG	TIFWN	OVPVI	KAFT	YKNE	F	FTDTVT	-DNGT	KKTC	GIPKV
GNNFPDGG	E <mark>E</mark> I	<b>Y</b> AFQSKM	V I GRVI	DAAGRK	LAV	RKSRNGG	LVLAG	< SR E K <mark>Y</mark> I	NSDKNR	YVNV	F I	SSDTMVV	-GNGA	F DWN <mark>M</mark> C F	F PWHI
	Q <mark>ER</mark>	WVDDDFY	LFTRG	P – D E K <mark>V</mark>	VVAL	TRNGSGSQ <sup>-</sup>	LQRT	PNLPH	ND <mark>G</mark> TI	<b>LTNI</b>	F [	P S DQ T C	:-TQ <mark>G</mark> K	TVSTN	GQPK I
- SLGQS <mark>FI</mark> DSA-C	T <mark>E</mark> S	LYDSDFY	SFTRH	E T	YLAT	TNRGSGAS-	GIQRT	SDSGF	SD <mark>G</mark> T'	VL SNV	L I	PSDRVTV	-DDGK	FTVTLSK	GEPK
S L G Q F F I NS E - Q V L G T S F T A A D - Q S L G Q N F D S D - Q			SEIR				VS	2111PK	DSDFV	LANV		PSDKVIV			GEPKI
			CETDIN	)/		IK GEFGN	LSKK				<mark> </mark>		-9516	DISTIC	GEPKL
- SL GONE DSD-C	<del>i</del> Et	TYDDEFY	SETRI	бт	VAL	SNKGSOTV-	-VERT	ТбнбЕ	SNGD		i I		- ENDE		GEPK
- LLGANFLNAD - C	IER	NQDDSFF	AFTRH	)№	IL VC S	TN GDNNN-	LQRT	TYHDY	TVGT'		Y	DENDRVT	-DTSG	TINLIS	GEGK
- EQGSSLYEAA - C	VER	YVDDEFF	AFTRG	√ <mark></mark> √	(FVAT	SNIGTGQS-	-LSRP	TYHPY	SS <mark>G</mark> T'	TLVN	LI	DSSDKVTV	- ENDQ	FEVNIED	GMPKV
- S L G S Q L Y NAQ - <mark>Q</mark>	I <mark>ER</mark>	YVDDQF <mark>F</mark>	AF S <mark>R</mark> G	√ <mark></mark>	LVAT	TNLGSGQS-	- FTR <mark>T</mark>	TYHPY	SD <mark>G</mark> TI	< LVNQ	L I		-N <mark>NG</mark> Q	FQ <mark>V</mark> N <mark>I</mark> NN	
-SVGSALYESD-G	VQR	YVDDQFF	AFTRG	<u> </u>	FVAT	TNLGSGQS	FTRT	TYHPY	SDGTI	VEVNQ	L	DTSDSVV	-NNGQ	FQVYKN	GLPK
-SRGSDLYEAP-C	VER VER	YVCNDCFF	AFIRG		FVAI	SN GGGQS	TCKT								GMPVV
	VER		AESPG	) <u>N</u>		CODATVH-							SCDCP		INAVO
	VER	HVEPDVY	AESRG	É <mark>v</mark>	1 VVV	CODATVH	·   T	DDHPY	OPGDI	TTNV		TRETER	SGDGK		GEPI I
- ALGKEWLREK-C	VER	HVESDVY	AFSRG	< <mark> </mark>		TONTTVK-	IK	TNSPY	EPGN'	TLKNV	L I	VPTEKFKV	SLDGS	LPVTLNL	GOPLV
- AQGDSWLKSD-C	I <mark>ER</mark>	HVSPNVY	SESRDI	< <mark> </mark>	LIVE	TSSNTTT	AK	ESHPY	KA <mark>G</mark> Q(		L I	VQK QT F K	SSTGH	LVATLYD	GEPLV
- SLGNEWIHEK-C	MER	HVQTDAY	AFSKG:	SA		TNLDTS IN-	IS	RDTPY	QP <mark>G</mark> D'	TLGN	L I	VP SETLT	SSDGS	LSVTLSS	GQPLV
- SQGK SWLNSN-C	LER	HVDAKVY	AFSRO	ý <mark>v</mark>		TNSMS T	RV	KSHPF	KAGD	EVQNL	L	VKQKFK	SNEGQ	LKVTLLD	GEAL
	I <mark>ER</mark> I <mark>E</mark> R	HVSPNVY VVTDNVV	SESRE			TANETTT						ICK OME K	AENGE		GEPLV
	VTWERYLLPLVER	HVOSNAY	AFSRG		MVV		AN T	FHSPY	KPGT				SSDSS		GEPLV
	VER														
- SLGEDWLOOK -C	VER	HVOSDVY	AESKG	) <mark> </mark>	MVVV	ALERTAN-	·   T	ESSPY	EPGN	<b>I</b> RNV	<u> </u>	VP S DV F R	SSDGS	PVTENS	GKPL L
- S L GQ P <mark>W I</mark> S S P - <mark>C</mark>	I EQ	HVEPEVY	AF SR RI	< <mark>\</mark>		TRVTTAS	• T <mark>T</mark>	QSHPF	NE <mark>G</mark> EI		L/	MTQTFL	DAKGR	LEAKMVS	GEPLV
- AL GNKWLQEK - C	VER	HVQPDAY	AFSRG	<	LVV	TOTTTVE	IK	RNSPY	EPGN	TLRNV	L [	RTETFK	SSDRS	LLVTLNL	GQPVV
-ALGKEWLQEE-C	VER IER	HVESDVY	AFSRG	< <mark>!</mark>			·IK	INSPY				PMEKFK	SPUGS		GOPLV
	I <mark>ER</mark>														
- AMRAGWLKSP-C	IER	HVDGEVV	AESROI	< <mark> </mark>		TTESKNTT-	RT	OSHPE	EADE	V F NI		TTOKET	SKEGK		GEPLV
-KFPKSWLKSP-C	IEO	YVS SNVY	AFSRD	ЕТ	LTF	TTEKKE IT-	VI.	TKSHPF	SAGD'		L I	VPTHTTK	SKENE	LKLTMYG	GEPVI
-NITEDLT AQ	VQR	Y SDNGFY	AFTRG	N∀	FVAL	TNGGSNEV-	OTSRN	TYHPY	LE <mark>G</mark> T'	TLCN	FYI	PTEDCIKN	-AGGA	F T <mark>V</mark> YLDH	GESK
-VKQQSLYQYP-C	I <mark>QR</mark>	YADDSFY	AFTRG	2T	FVAL	TNGGSNQG-	QISRT	TYQPY	QDGT'	TLCN	FWI	PTEDCIK	-VDGQ	FTTYLDN	GESK
	VQR LER														
	ŸEK			5 <mark>X</mark>		INGGCGAO	-TGAS				í		-OGGT		CNIDKV
-AVNAGAOFC	VQR	YADNDEY	AFTRG	) <mark>i</mark>	FAAF	TSKYDRDV	IRK	TYHPY	ADGTI	<b>CNV</b>		TGDCVIN	-ONGS	FOVYERN	GEVK
- ITNAGLQGG	T <mark>E</mark> A	YCDQDLY	AF F RGI	( <mark>L</mark>	FVGV	TSK FNQVT-	RQ	VTHPF	SN <mark>G</mark> E'	<b>VICNI</b>	F Y I	ΥΤΟ <mark>DCL</mark> Τ	-TNNS	FNLVLIG	GEAK
KSK WNOPW	VER	Y SNPNFY	SESRG	E – – – F	LVAL	TN SNDOOH-	FK	TYHPE	KD <mark>G</mark> Q'	<b>CN</b>	FYI	PTSDCQTV	NGG	VOVYLLN	GESK
	VER														
-TUPOFFLNND	VQR VER	T S I DINFY	AFSRG	5T		INSGASED		TYUNY			LAVP		GASGA	FINSMIGK	
	NP I														
SKDTGEVTYK-N	WP 1	KDDTT	AMRKG	TDGSO	VTIL	SNKGASGD	SYTIS	SGAGY	TAGO0	TFM		GCTTVT	GSDGN	VPVPMAG	GI PRV
SADSAY TYA-N	DAF	YTD SNT I	AMRKG	TS GSQ <mark>V</mark>	TVL	SNKGSSGS-	SYTLT	SGSGY	TSGTI	<liea< td=""><td>Y ·</td><td>TCTSVTV</td><td>DSSGD</td><td>I PVPMAS</td><td>GLPRV</td></liea<>	Y ·	TCTSVTV	DSSGD	I PVPMAS	GLPRV
-AVGGLGGNDH	KHL	Y SQD S A Y	AWSRAI	DGDL	VLT	L <mark>N</mark> RGQGYS-	GQ'	YCFNTG	KNNK'	TWDR	F	-GSGT <mark>V</mark> TS	DG <mark>NG</mark> Q	VC <mark>V</mark> S YTN	GEPEV
AVDAT <mark>YL</mark> TYK - N	WV I	Y SDTTT I	AMR <mark>K</mark> G	YDGKQ <mark>V</mark>	TVL	SNKGASGD	AYTLN	SNTG	TS <mark>G</mark> TI	E <mark>V</mark> VE	L ·	TCATT	TSGNT	VTVPMAN	GLPR
SQDDGYL SYK-S	NP I	YSDSHTI	AMRKG	2P GAQ	GVF	INVGTFSS/	AEATLP	SATGE	EAGQ/		<u></u>	GCTEHTT	DSGGG	AVTLAS	GLPVV
VASANSE <mark>EY</mark> AFP=V	KFP FVT	VSSISQM	AVSKPI	<mark> </mark>	LALL	INQGSSSS-	- PQWT	GSAGY	DANE		<b>v</b>	-SCIK <mark>I</mark> TA	UG <mark>NG</mark> G	V 5 I I AQC	GKPMV