FEBS Letters 581 (2007) 3927–3935

Where do animal α-amylases come from? An interkingdom trip

Jean-Luc Da Lage[a,b,*], Etienne G.J. Danchin[c], Didier Casane[a,b]

[a] *Laboratoire Evolution, génomes et spéciation (LEGS), CNRS, 91198 Gif sur Yvette cedex, France*
[b] *Université Paris-Sud 11, 91405 Orsay cedex, France*
[c] *AFMB-UMR 6098, CNRS, Universités Aix-Marseille I et II, Case 932, 163 Avenue de Luminy, 13288 Marseille cedex 09, France*

Received 11 May 2007; revised 3 July 2007; accepted 6 July 2007

Available online 20 July 2007

Edited by Takashi Gojobori

Abstract **Alpha-amylases are widely found in eukaryotes and prokaryotes. Few amino acids are conserved among these organisms, but at an intra-kingdom level, conserved protein domains exist. In animals, numerous conserved stretches are considered as typical of animal α-amylases. Searching databases, we found no animal-type α-amylases outside the Bilateria. Instead, we found in the sponge *Reniera* sp. and in the sea anemone *Nematostella vectensis*, α-amylases whose most similar cognate was that of the amoeba *Dictyostelium discoideum*. We found that this "Dictyo-type" α-amylase was shared not only by these non-Bilaterian animals, but also by other Amoebozoa, Choanoflagellates, and Fungi. This suggested that the Dictyo-type α-amylase was present in the last common ancestor of Unikonts. The additional presence of the Dictyo-type in some Ciliates and Excavates, suggests that horizontal gene transfers may have occurred among Eukaryotes. We have also detected putative interkingdom transfers of amylase genes, which obscured the historical reconstitution. Several alternative scenarii are discussed.**

Keywords: Amylase; Lateral gene transfer; Unikonts; Cnidaria; Porifera; Intron gain

1. Introduction

Most living organisms are able to exploit environmental polysaccharides. Alpha-amylases (EC 3.2.1.1), which hydrolyze starch and other polysaccharides in maltose, maltotriose and maltodextrines, are key enzymes in this process. They are widespread in animals, fungi, plants, and are also found in the unicellular eukaryotes, Eubacteria and Archaea. However, they may have been lost in some parasitic organisms [1]. In many cases, they form multigene families (e.g., [2–6]). The presence of several, sometimes divergent amylase molecules enables organisms to digest a broad range of substrates, in a broad range of environmental conditions. The tridimensional structure of the enzyme is well conserved and about 10 amino acids are conserved in known α-amylase sequences, three of which, two aspartic acids and one glutamic acid, form the "catalytic triad" [7–11]. Glycosyl hydrolases (GH) have been classified into different families based on sequence similarity and common fold [12,13] (CAZy database http://www.cazy.org/index.html). Alpha-amylases are found in families GH13 and GH57. GH13 contains most α-amylases, and related enzymes that cover at least 26 different EC numbers. This family has been subdivided into 35 subfamilies (Fig. 1, [14]). As a whole, five conserved short amino acid stretches have been reported [7,15]. However, at the intra-kingdom level, there is much greater sequence similarity and homogeneity, except in the Bacteria, where different types coexist [11,14]. In animals, the overall sequence identity between amylases identified until now is over 40%, and typical motifs have been described [11,16]. Amylase genes have been cloned from a number of Protostomes and Deuterostomes (subfamilies GH13_15 and 24, respectively [14]). However, until now, no data on amylases in non-Bilaterian phyla, such as Porifera (sponges) and Cnidaria, were available.

Numerous genome data are now available, and we have screened non-Bilaterian genomes for animal-type amylases. Surprisingly, we failed to find any. However, other putative α-amylase sequences were found, with high similarity to that of the slime mold *Dictyostelium discoideum* (Amoebozoa, Mycetozoa). In this paper, we show that the "Dictyo-type" α-amylase is widespread and may be ancestral in the Unikonts, a clade comprising Animals, Fungi (Opisthokonts) and Amoebozoa.

2. Material and methods

Table 1 shows the sequences analyzed in this study with their accession or code number. URLs of the online databases are detailed in the legend. Table 1 lists the species of interest to our study, but many more species for which genome data were available in the mentioned databases were screened as described below. This includes for instance all Prokaryotes and Protists genomes in Entrez Genome (NCBI).

Traces archives (the raw, uncorrected sequencing runs) of *Nematostella vectensis* (Cnidaria, Anthozoa), *Hydra magnipapillata* (Cnidaria, Hydrozoa), and *Reniera* sp. (Porifera, Demospongiae), were downloaded and searched by TBLASTN [17], using the amylase sequence of the Bivalve *Corbicula fluminea* (AF468016) as a query for animal-type amylase (subfamily GH13_15). The *Aspergillus niger* (Ascomycetes) amylase (A35282) was used as a fungus-type (GH13_1); a barley amylase (P04063) was used as plant-type (GH13_6). In the case of significant hits, a protein sequence as complete as possible was reconstructed from the partial subject sequences. Reciprocal BLASTP was then performed against GenBank using the reconstituted putative amylase protein. This protocol was also used for the other species available as traces archives (Table 1). The searches were not filtered for low complexity regions and default parameters were used. When *Nematostella* contigs were released at the Joint Genome Institut (JGI), these were screened online by TBLASTN to improve and

*Corresponding author. Address: Laboratoire Evolution, génomes et spéciation (LEGS), CNRS, 91198 Gif sur Yvette cedex, France. Fax: +33 1 69 82 37 36.
E-mail address: jldl@legs.cnrs-gif.fr (J.-L. Da Lage).

0014-5793/$32.00
doi:10.1016/j.febslet.2007.07.019

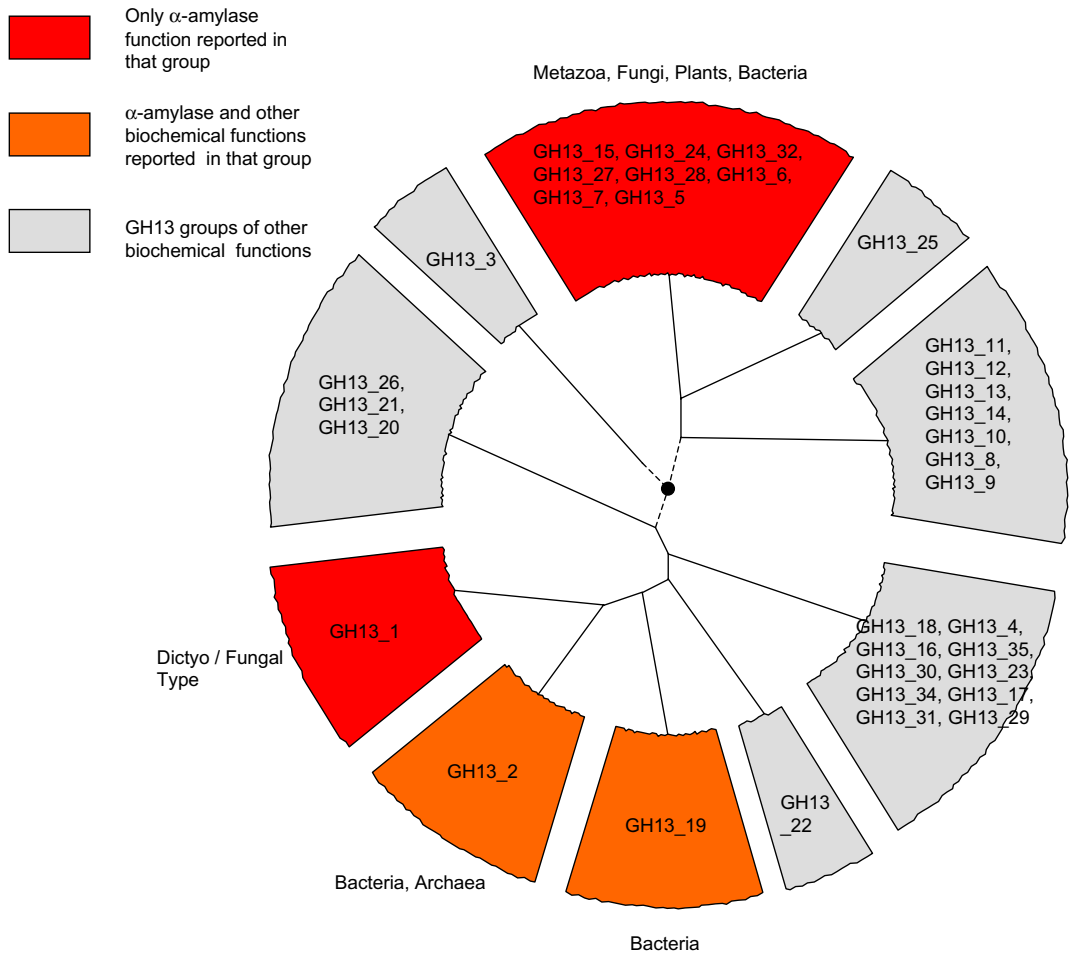

Fig. 1. Simplified classification of the glycosyl hydrolases of the GH13 family (after Ref. [14]).

complete our first sequence. Since the first results have shown the interest of the GH13_1 *Dictyostelium discoideum* α-amylase (**XP_640516**), this sequence was also used as a query in BLAST searches.

When possible, e.g. for *Reniera*, intron positions were inferred by comparing the reconstructed protein sequence with the DNA sequence from traces data. For some species, intron data were available; for others, only ESTs were found. The retrieved putative amylase protein sequences of the *Dictyostelium* type were aligned with ClustalW [18] using default parameters. A tree was built from the protein alignment, after removing ambiguous positions, i.e. on 266 remaining positions (alignment as supplementary data 1). After having chosen the appropriate model with PROTTEST1.3 [19], we used maximum likelihood (PHYML [20] with parameters WAG + Γ + I), Bayesian inference (BI, MrBayes [21] with parameters WAG + Γ + I) and neighbor joining (made using MEGA [22], Poisson correction + Γ, $\alpha = 1$ estimated by PHYML). All methods gave essentially congruent topologies.

3. Results

3.1. Search in non-Bilateria metazoa

No animal-type amylases were found in the non-Bilateria genomes examined. Screening traces and online databases, including the Cnidaria database CnidBase (cnidbase.bu.edu/), with the *Corbicula* sequence resulted in no hits. We searched the same databases with a fungal-type sequence, *A. niger*. Many hits were found in *Nematostella* traces, with *E*-values ranging from 10^{-10} to 10^{-51} and sequence lengths around 800–1000 bp. In *Reniera*, significant values were also obtained (10^{-10}–10^{-28}). In contrast, there were no hits using the same query with *Hydra* traces. For *Nematostella* and *Reniera*, reciprocal BLASTP in GenBank returned the *Dictyostelium discoideum* putative α-amylase (GH13_1) as the best score, with *E*-values of 10^{-60} and 10^{-54}, respectively. We also screened the traces with a plant-type α-amylase (GH13_6, *Hordeum* sequence). For *Nematostella* traces, we obtained a few significant hits, with *E*-values up to 10^{-26}. The reciprocal BLASTP in GenBank showed strong similarity with Archaeal α-amylases from *Thermococcus* or *Pyrococcus* (GH13_7). The traces were therefore screened again with a *Thermococcus* amylase sequence (AAC97877), with *E*-values reaching 10^{-89}. The reconstructed *Nematostella* sequence was used as a query against GenBank, the best hit found was an *E*-value of 10^{-123} with *P. furiosus* (AF001268). In *Reniera*, no similarity was found with the *Thermococcus* sequence.

The above results could be suggestive of contaminations of genomic DNAs by exogenous DNA, such as symbiotic protists or other associated microorganisms, since Cnidaria are known to harbor many organisms [23,24]. We checked for archaeal DNA contamination by screening the *Nematostella* traces with the DNA polymerase Pfu, a typical *P. furiosus* archaeal gene. The result was negative. For *Nematostella*, genome scaffolds have now been publicly released at the JGI. We confirmed

Table 1
The species mentioned in this article are listed by alphabetical order

Species name (Taxonomy)	Genome database used	Release or date	Accession or data reference
Acanthamoeba castellanii (Amoebozoa)	Traces NCBI, HGSC Baylor College of Medicine	October 2006	e.g. D9C6L5U01AGTUI, D5NSVWU02FI8Q5
Aspergillus fumigatus (Ascomycetes)	GenBank	April 2006	XP_751813
Aspergillus niger (Ascomycetes)	GenBank	September 1999	A35282
Aspergillus oryzae (Ascomycetes)	GenBank	April 1993	AAA32708
Chaetomium globosum (Ascomycetes)	GenBank, Broad Institute (MIT)	November 2006	supercontig 1.4, 3861549 3862802
Coprinopsis cinerea (Basidiomycetes)	Broad Institute (MIT)	v1.0	contig 90, 26400–27272
Debaryomyces occidentalis (Ascomycetes)	GenBank	April 2005	CAA43995
Dictyostelium discoideum (Mycetozoa)	GenBank	August 2006	XP_640516
Entamoeba histolytica (Amoebozoa)	Traces NCBI, Sanger Centre, GenBank	January 2007	XP_651807
Fusicoccum sp. (Ascomycetes)	GenBank	July 2006	ABG48762
Giberella moniliformis (Ascomycetes)	GenBank	August 2005	AAZ73168
Hartmanella vermiformis (Amoebozoa)	TBestDB (Anabench)	April 2004	HVL00001159
Herpetosiphon aurantiacus (green non-sulfur bacterium)	GenBank	January 2007	EAU16952
Histiona aroides (Excavates, Jakobids)	TBestDB (Anabench)	July 2006	HAL00000515
Hydra magnipapillata (Cnidaria)	Traces NCBI, GenBank, Hydra EST database	April 2006	Not found
Jakoba libera (Excavates, Jakobids)	TBestDB (Anabench)	July 2006	JLL00001190
Laccaria bicolor (Basidiomycetes)	JGI	v1.0	scaffold_57, 58609-60711
Lipomyces starkeyi (Ascomycetes)	GenBank	September 2006	AAN75021
Malawimonas californiana (Excavates, Malawimonadids)	TBestDB (Anabench)	July 2006	MCL00001248
Mastigamoeba balamuthi (Amoebozoa)	GenBank, TbestDB (Anabench)	jul. 2006	ABD46606 (GenBank), MBL00002638 (TbestDB)
Monosiga brevicollis (Choanoflagellates)	Traces NCBI, JGI	v1.0	scaffold_2, 655806 658210
Mycosphaerella graminicola (Ascomycetes)	JGI	v1.0	scaffold_10, 450481 452082
Naegleria gruberi (Excavates, Heterolobosea)	JGI	v1.0	scaffold_28, 239206 240736
Nematostella vectensis (Cnidaria)	Traces NCBI, JGI, CnidBase	v1.0	scaffold_131, 530958 533591
Nyctotherus ovalis (Ciliates)	GenBank	November 2006	CAI59813
Oxytricha trifallax (Ciliates)	Traces NCBI	February 2007	e.g. OXAO aas44e03.b1, OXAO aak83g06.g1
Paramecium tetraurelia (Ciliates)	ParameciumDB	v1.0	GSPATT00010345001
Phakopsora pachyrhizi (Basidiomycetes)	Traces NCBI	October 2006	Not found
Phanerochaete chrysosporium (Basidiomycetes)	JGI	v2.0	scaffold_12, 835115 836624
Physarum polycephalum (Mycetozoa)	GenBank	December 2006	ABD46585
Reclinomonas americana (Excavates, Jakobids)	TBestDB (Anabench)	jul. 2006	RAL00004060
Reniera sp. (Porifera)	Traces NCBI	February 2006	BAYB423511.b1, BAYB105355.g1
Saccharomycopsis fibuligera (Ascomycetes)	GenBank	April 2005	P21567
Schistosoma mansoni (Platyhelminthes)	TIGR	January 2007	Not found
Schmidtea mediterranea (Platyhelminthes)	Traces NCBI, WU GSC	v3.1	Not found
Seculamonas ecuadoriensis (Excavates, Jakobids)	TBestDB (Anabench)	July 2006	SEL00001840
Tetrahymena thermophila (Ciliates)	Tetrahymena genome database, GenBank	November 2006	EAS00610
Thermoactinomyces vulgaris (Eubacteria, Firmicutes)	GenBank	April 2005	CAA49465
Trichoderma reesi (Ascomycetes)	JGI	v2.0	scaffold_6, 583598 585145
Trichomonas vaginalis (Excavates, Parabasalids)	GenBank	January 2007	EAY23285
Trichoplax adhaerens (Placozoa)	Traces NCBI	November 2005	Not found
Trimastix pyriformis (Excavates, Metamonads)	TBestDB (Anabench)	July 2006	TPL00000301

The databases used for their study are indicated and the accession numbers or sequence references are given. Internet addresses of the databases: Traces NCBI: www.ncbi.nlm.nih.gov/Traces/; GenBank: www.ncbi.nlm.nih.gov/entrez/; Baylor College of Medicine: www.hgsc.bcm.tmc.edu/; TBestDB: tbestdb.bcm.umontreal.ca/searches/welcome.php; Broad Institute: www.broad.mit.edu/tools/data/seq.html; Sanger Centre: www.sanger.ac.uk/DataSearch/blast.shtml; Hydra EST database: hydra1.calit2.uci.edu/hydra-server1/blast/; CnidBase: http://cnidbase.bu.edu/; JGI (Joint Genome Institute): genome.jgi-psf.org/; Paramecium DB: paramecium.cgm.cnrs-gif.fr/db/index; TIGR: tigrblast.tigr.org/tgi/; WU-GSC: genome.wustl.edu/tools/blast/; Tetrahymena genome database: www.ciliate.org/.

the presence of the Dictyo-type sequence traces (accession in Table 1). More importantly, we have checked that this gene is surrounded by *bona fide* metazoan genes, and not by *Dictyostelium*-like sequences: the neighboring genes are highly similar to for example, human inositol triose-phosphate receptor ($E = 0.0$), Xenope Rg9mtd2 ($E = 10^{-56}$), and *Haliotis* macrophage expressed protein ($E = 10^{-141}$). We are thus confident that the Dictyo-type α-amylase gene found in *Nematostella* is not an artifact. In *Reniera*, whose genome has not been assembled yet, we have screened the traces with two *Dictyostelium* genes, an RNA polymerase and a CaM-phosphatase. The result was negative. We concluded that these two species have an α-amylase gene highly similar to that of *Dictyostelium*.

The question of an Archaeal-type amylase in *Nematostella* is less clear. The very strong similarity of the sequence found in the traces with the *P. furiosus* α-amylase makes it suspect. In the annotated database at the JGI, a BLASTP search using *P. furiosus* gave no result. A TBLASTN showed a significant

hit, in a very short scaffold 7904 (3280 bp), which suggests archaeal DNA contamination.

We found no sequence that was clearly an amylase sequence of any type in the Hydrozoa *Hydra magnipapillata*, except for a few sequences that are clearly bacterial contamination (not shown).

Few traces archives were available for the simple Metazoan *Trichoplax adhaerens* (Placozoa), and the BLAST search was unfruitful. We also failed to find any α-amylase in the Bilaterian Platyhelminths worms *Schmidtea mediterranea* and *Schistosoma mansoni*, except for bacterial contamination (not shown).

3.2. Search in relatives of metazoa

Based on the above observations, we hypothesized that Dictyo-type amylases could be ancestral in animals. Therefore, we screened other organisms for similar amylase genes, and for animal-like amylases. Except when mentioned, no animal-type amylases were found. We first screened the Choanoflagellate *Monosiga brevicollis* traces. Choanoflagellates are a phylum considered a sister group to Metazoans [25,26]. Using *A. niger* as a query, we obtained *E*-values of 10^{-28}. The reciprocal BLASTP against GenBank showed high similarity with the *Dictyostelium* α-amylase. The *M. brevicollis* genome has been released at the JGI, we were therefore able to confirm our initial results: we have obtained a full amylase sequence, using BLASTP and the *Dictyostelium* amylase as a query ($E = 10^{-122}$). Thus, *M. brevicollis* has an amylase of the GH13_1 Dictyo-type.

3.3. Search in fungi

Since reciprocal BLAST searches always showed *Dictyostelium* as the best hit when *A. niger* was used in the initial search, it was not surprising that there was a strong similarity between the amylase sequences of these organisms. Using the *Dictyostelium* α-amylase as a query in GenBank (BLASTP), significant hits were found with Fungi (e.g. *Aspergillus terreus*, $E = 10^{-95}$). All could be classified in the GH13_1 subfamily, along with *Dictyostelium* α-amylase [14] (for our study, Ascomycetes and Basidiomycetes were sampled in databases, see Table 1). In Fungi, animal-type amylases were absent from the species studied (e.g. at the Broad Institute and CAZy databases), except in traces of *Phakopsora pachyrhizi*, where a contamination by dipteran DNA was likely (not shown), and in *Phanerochaete chrysosporium* (Basidiomycete) and *Chaetomium globosum* (Ascomycete) (Table 1).

These last two cases deserve more detailed explanations: the *E*-values with an animal (*Corbicula*) query were 10^{-67} and 0.0, respectively. For both species, the reciprocal search has shown the best hits to be Actinomycetales (10^{-140} and 0.0, respectively), which are Actinobacteria. For example, there was 73% identity and 84% similarity between the *C. globosum* and the *Streptomyces coelicolor* (**NP_631084**) α-amylases. Since the sequences were found inside large-size genomic contigs, we are confident that they are not due to contaminating material. Streptomycete α-amylases are known to be close to animal-type amylases [11,14,27]. These two fungi may have gained an α-amylase gene each, probably separately, through lateral gene transfer from Actinobacteria. The *P. chrysosporium* protein has an extra C-terminal starch binding domain (SBD) of the CBM20 family, which is frequent in bacterial α-amylases, including in Actinomycetales (GH13_32), but is more rare in Eukaryotes, and has not been found to date in animal amylases [28,29]. However, since SBDs have also been found in some fungal α-amylases (CAZy database), it is not clear whether the *P. chrysosporium* SBD has been transferred from an Actinobacterium, along with the core enzyme or if it has been grafted from an endogenous sequence. Another striking observation, given this likely transfer from Prokaryotes to Eukaryotes, is that the *P. chrysosporium* coding gene (scaffold 12, 835019 837355) has 11 introns, which have thus all been gained recently. There are no introns in the *C. globosum* animal-type amylase. Also note that these two fungi have also a GH13_1 amylase (Fig. 3).

3.4. Search in amoebozoa

We also screened a number of protist genomes and EST databases, starting with relatives of *Dictyostelium* itself, the Amoebozoa [30–32]. Again, no typical animal-type amylase was found in any protist species available. We found Dictyo-type amylase sequences in *Physarum polycephalum* ($E = 10^{-57}$, a moderate value due to a partial sequence), *Acanthamoeba castellanii* ($E = 10^{-129}$, the best hit (*Dictyostelium*) of the reciprocal BLASTP against GenBank after reconstruction of the sequence of *A. castellanii*, for which only very small size traces sequences were available), *Mastigamoeba balamuthi* ($E = 10^{-91}$), *Hartmanella vermiformis* ($E = 10^{-68}$, partial EST sequence).

We found no Dictyo-type α-amylases in *Entamoeba histolytica* traces. However, we found significant hits ($E = 10^{-28}$, 48% of positives over 310 residues) using animal-type amylase as a query. This result corresponded to a protein distantly related to α-amylase (**XP_651807**). This sequence was also highly similar to two Basidiomycete *Coprinopsis cinerea* genes CC1G_08568 and CC1G_05516, and to a *Trichomonas* gene (**EAY23285**). High similarity was also found in the Ciliates *Tetrahymena* (EAR85737), *Paramecium* (Scaffold 46, 241098–241803) and *Oxytricha* (trace OXAO aat84g02.b1). Although less similar, but expected since it was the initial query, animal α-amylases were also related to this sequence (*Diabrotica virgifera* α-amylase, $E = 10^{-32}$). Unlike *Coprinopsis*, we found no similarity with this sequence in the genomes of two other Basidiomycetes, *Laccaria bicolor* and *Phanerochaete chrysosporium*.

3.5. Search in other protists

Using BLASTP or TBLASTN with the α-amylase of *Dictyostelium* as a query, we have also found strong similarities in other genomes, sometimes very distant from this species and from Unikonts as a whole. Similar sequences were recovered from several Jakobidae species (clade Excavates), e.g. *Jakoba libera* ($E = 10^{-76}$), in *Malawimonas californiana* (Excavate, $E = 10^{-69}$), in *Trimastix pyriformis* (Excavate, $E = 10^{-136}$), in *Naegleria gruberi* (Excavate, $E = 10^{-143}$). Three Ciliates also harbored a Dictyo-type amylase. These were *Tetrahymena thermophila* ($E = 10^{-111}$), *Paramecium tetraurelia* ($E = 10^{-108}$), whose genome assembly is available, and *Oxytricha trifallax*, whose traces only were available (best reciprocal BLASTP hit: *Dictyostelium*, $E = 10^{-101}$). A partial sequence annotated in GenBank as α-glucosidase of *Nyctotherus ovalis* (Ciliate) was also recovered with a significant *E* value (10^{-53}, length of 187 amino acids). It is likely to be a GH13_1 α-amylase.

3.6. Search in green plants

We found no animal-type nor Dictyo-type α-amylases in plant genomes and sequences available at the NCBI. In plants, all reported α-amylase activities pertain to subfamily GH13_6.

3.7. Search in bacteria

A few examples of animal-type amylases have been identified in various, unrelated bacterial lineages [11,27,33–35]. To get a more complete view of the distribution of Dictyo-type amylases in living organisms, we searched the Entrez Bacterial genomes, (718 species) and GenBank nr (Bacteria) databases with TBLASTN, using the *Aspergillus niger* α-amylase as a query. We found a few hits with *E* values around 10^{-61}, e.g. the Firmicute (Bacillales) *Thermoactinomyces vulgaris* and the Green non-sulfur bacterium *Herpetosiphon aurantiacus* (Chloroflexi). Interestingly, the reciprocal BLASTP gave *Dictyostelium* as the best eukaryote hit. However, those sequences are too distant to be considered as GH13_1. Lateral transfers from Unikonts are possible but not likely given the current data.

A consensus phylogenetic tree of Eukaryotes (Fig. 2) summarizes the positions of the species which possess an amylase of the fungus/Dictyo-type. It appears that this type is widely distributed in Unikonts, except in *Entamoeba* and in bilaterian animals (Protostomes and Deuterostomes). In addition, such α-amylases are present in several, unrelated clades of the Bikonts. We have tried to clarify the relationships among the sequences in order to distinguish between common ancestry and lateral transfer. Dictyo-type amylase sequences for which

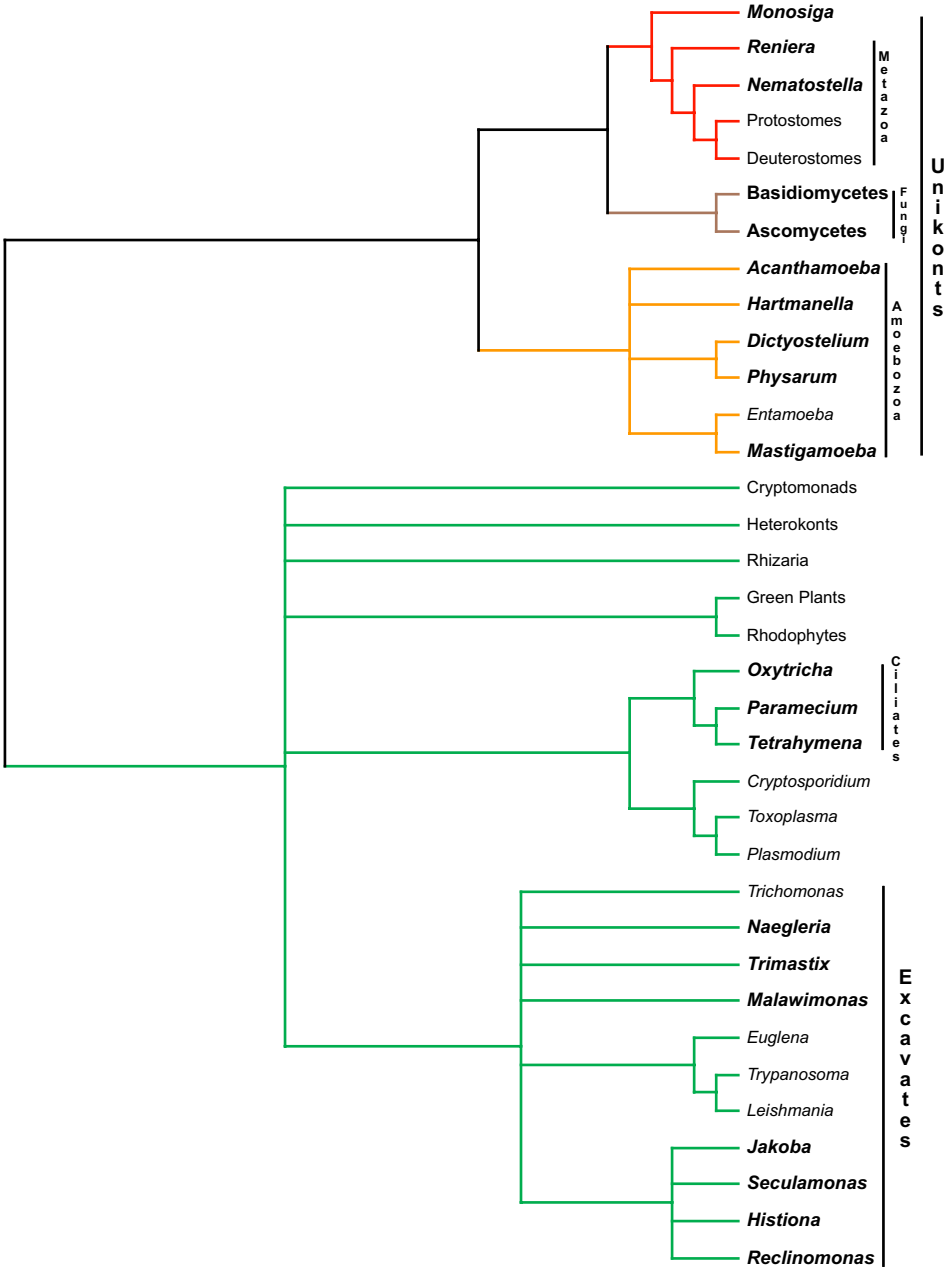

Fig. 2. Phylogenetic tree of eukaryotes made by consensus of several published molecular phylogenies [30,32,36,50–52]. The organisms or clades which possess an α-amylase of the Dictyo-type are in bold type. Fungi are colored in beige; animals are red; amoebozoa are orange; bikonts are green.

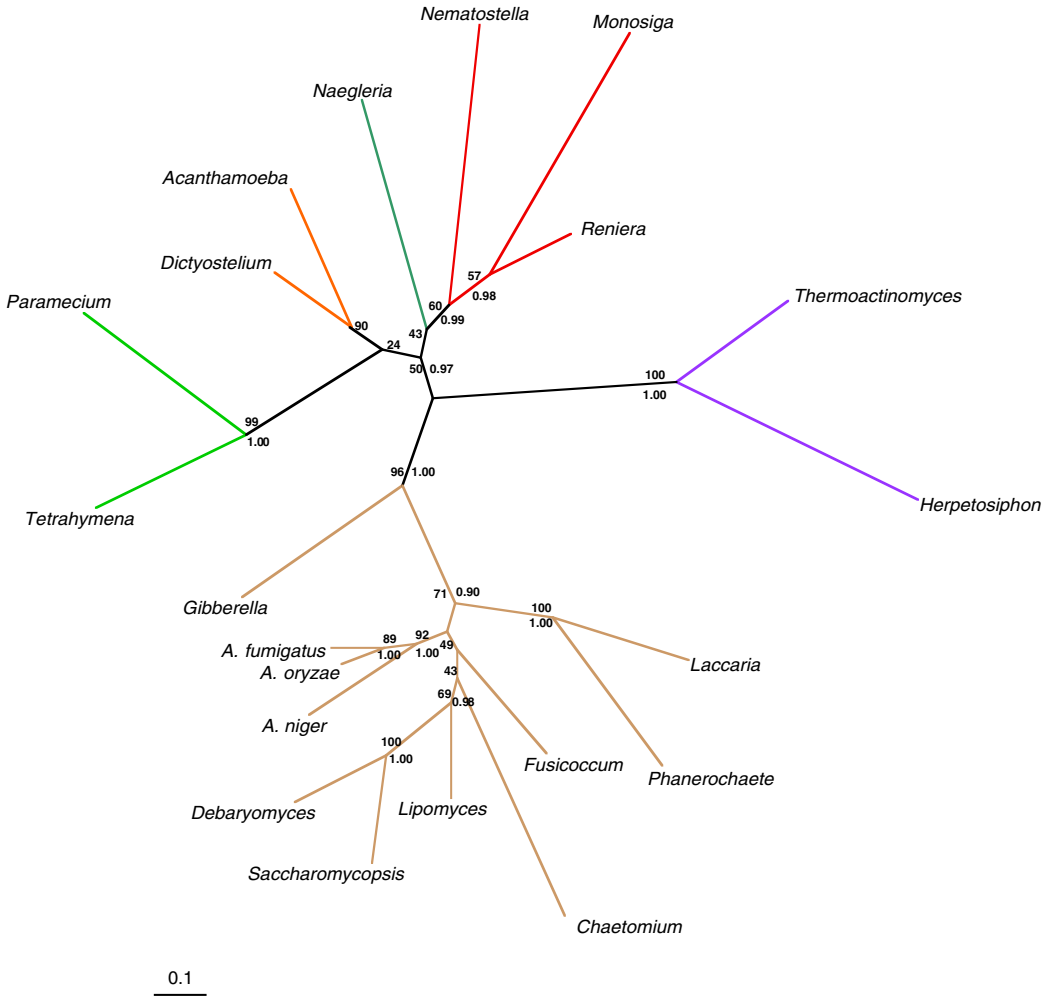

Fig. 3. Radial ML tree of the α-amylase sequences of the Dictyo-type, for which the longest sequences were available (266 positions retained). Numbers at nodes are bootstrap values (above or left to branches) and posterior probabilities from the BI tree (below or right to branches, if >0.50). Scale bar shows the substitution number per site. Fungi are colored in beige; animals are red; amoebozoa are orange; bikonts are green; bacteria are purple.

we have enough sequence length were aligned (Supplementary material 1), and a gene tree was inferred (Fig. 3). All sequences are from the subfamily GH13_1, except the two Bacterial sequences. Sequences from the sponge *Reniera* and the Choanoflagellate *Monosiga* clearly clustered together, and then with the cnidarian *Nematostella*. The sequences that were most closely related to them, were *Naegleria*, then the Amoebozoa *Dictyostelium* and *Acanthamoeba*, along with the Ciliates sequences. The Fungi formed a clearly separate group. This suggests that horizontal transfers have occurred in some groups, probably from Amoebozoa to some Excavates, and to some Ciliates. A tree inferred from shorter sequences (221 residues) but with more taxa, and less supported nodes (not shown), indicated that *Trimastix* and *Seculamonas*, two other Excavates, branched with *Naegleria*, and that the Amoebozoa *Hartmanella* branched with *Dictyostelium* and *Acanthamoeba*. Thus, if these branchings are to be confirmed as longer sequences become available, we may assume an old acquisition of this amylase gene in Excavates, and a firm ancestry in Unikonts.

An alignment of the Dictyo-type α-amylases with intron information is given as Supplementary material 2. Few intron positions are shared among these species, except between the two Basidiomycetes, and one position between *Reniera* and *Monosiga*.

4. Discussion

It should be first noted that, in many instances, the assignment of sequences as α-amylase was inferred by sequence similarity. Small amino acid changes may result in the modification of important enzymatic features [14]. It is therefore possible that some of the sequences studied here may have for example substrate specificities other than starch. We have shown here that the fungal/Dictyo-type α-amylase (i.e. family GH13_1) was probably ancestral (plesiomorphic) in Unikonts, based on the available data. The initial surprise was to find such sequences in a sea anemone and in a sponge. Obviously, the data are still scarce, and our conclusions would be strengthened by examining additional species. The problem of potential contamination requires careful attention, but our study allowed ruling out this hypothesis. The Dictyo-type amylase was not found in any bilaterian animals, which raises

interesting questions about the origin of animal amylases which are discussed below. On the other hand, Dictyo-type amylase was found in almost all Unikont species examined, except for *Entamoeba histolytica*. The absence of this type of amylase in *Entamoeba* could be related to its parasitic way of life. It has been reported that parasitic organisms are prone to gene losses, including energy and sugar metabolism associated genes [1]. However, little gene loss has been reported in the *Entamoeba* genome [31]. In addition, this species has several genes reported as α-amylases in GenBank, one of which appears to be rare in nature: we found similar sequences only in a few Ciliates, in *Trichomonas*, and in a fungus (*Coprinopsis*).

4.1. Origin of the Dictyo-type α-amylase

We assume that the Dictyo-type α-amylase was present in the ancestors of Unikonts. A major question, given our data, is whether this type is more ancestral. Is it ancestral to Eukaryotes as a whole? We have clearly identified this gene in the Ciliates, and in a number of Excavates. The first hypothesis is that it is not ancestral in Eukaryotes, but rather, that horizontal transfers occurred from non-fungal Unikonts. This is suggested by the positions of these bikont groups in the tree in Fig. 3. The relationships within Excavates are still unclear, in particular, the positions of *Trimastix* and *Malawimonas* are not firmly established [36]. If they are really related to *Naegleria* and to Jakobids, a single transfer event could account for the distribution found in this group. Another transfer event could have occurred to the Ciliates (Fig. 4A and B). A high frequency of lateral transfers has been reported among protists [37]. The fact that these organisms possess Dictyo-type, exogenous α-amylase genes, does not imply that they lack genuine α-amylases. For instance, as mentioned above, Ciliates have a different subfamily of amylases.

The alternative hypothesis is that the Dictyo-type is ancestral to Eukaryotes. This hypothesis cannot be ruled out with the current data and implies that it would have been lost several times, e.g. in plants, Heterokonts, Apicomplexa, Rhizaria, Euglenozoa (Figs. 2 and 4C). In this case, the GH13_1 like amylases found in *Herpetosiphon* and *Thermoactinomyces* could have originated from any Eukaryote harboring a GH13_1 gene (Fig. 4C). Note however that data are still scarce or totally missing for many of the Bikont taxa, except plants. Loss of the Dictyo-type would require the previous presence of another α-amylase in the genome, in order to maintain the function, except maybe in some parasitic organisms. Finally, this hypothesis seems currently less likely than horizontal transfers, because in case of descent from a common ancestral GH13_1 member, one would basically expect the Bikont sequences to cluster together, which is not the case (Fig. 3). However, the nodes are not strongly supported.

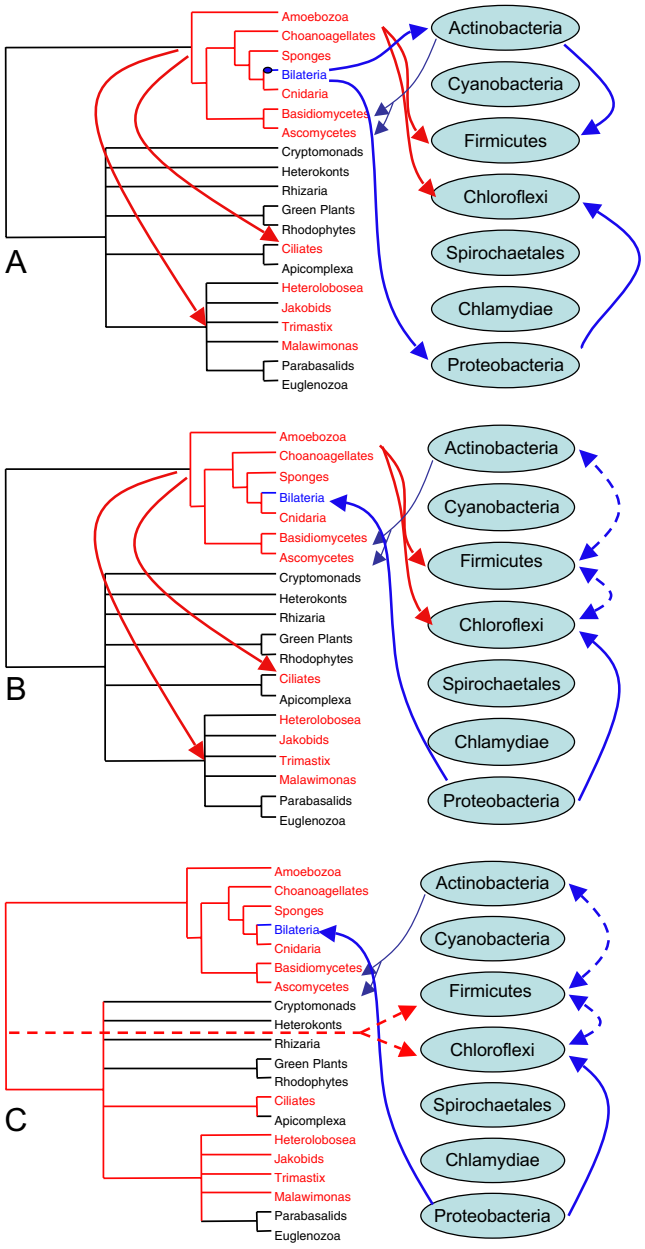

Fig. 4. Three scenarii about the origin of animal-type α-amylases. Red color is the Dictyo-type α-amylase; blue color is the animal-type; arrows indicate gene transfers. Dashed arrows indicate unclear, alternatively possible transfer events. Thin blue arrows are the back to eukaryotes transfers of animal-type amylases from *Streptomycetes* (GH13_32). (A) Intrinsic evolution from the Dictyo-type ancestral sequence (blue spot in the Bilateria branch). (B) Allogene origin through a transfer from a γ-proteobacterium related to Alteromonadaceae. (C) Variant of the B scenario, in which the Dictyo-type amylase is ancestral in eukaryotes.

4.2. Where do the animal amylase come from?

Various biological aspects of typical animal α-amylases (GH13_15, GH13_24) have been well studied, such as genetic variability [38–41], biochemistry, enzymology [42–47]. While the 3D structure is well conserved, the amino acid sequences are very different from those of the fungus/Dictyo-type. It is however possible to compare them by making a structural alignment, because of the conservation of the secondary structures (the $(\beta/\alpha)_8$ barrel). Between the human and the *A. niger* α-amylases, there is about 14% identity only, and 28% similarity (structural alignment made using DeepView [48]). We aim at explaining the origin of the Bilateria amylases, given an ancestral gene similar to the one found in non-Bilateria.

The first hypothesis is an intrinsic evolution from the ancestral type (Fig. 4A). However, divergence between the Dictyo-type (GH13_1) and the animal-type amylases (GH13 15/24) is about the highest among the GH13 family [14]. Thus, it

seems not realistic, except considering a very high evolutionary rate, followed by a slow down. Rather than GH13_1, an evolution from a GH13_5 ancestral sequence could be more plausible, since this subfamily is closer to the animal-type and is present in Fungi. However, there is no GH13_5 in non-Bilaterian animals. Based on the presence of animal-type α-amylases in a few dozen eubacterial species, we have previously hypothesized [11] several independent transfers from animals to Bacteria. If this is correct, considering our first hypothesis, the transfers should have occurred after the split of Cnidaria and the common ancestor of Bilateria (Fig. 4A). Later, a back to Eukaryotes transfer of an animal-type amylase (GH13_32) occurred twice, from Streptomycetes to some true fungi, probably accompanied, in one case, by a starch binding domain of bacterial origin. Note that we found few (and not ascertained) instances of transfers of Dictyo-type amylases to Bacteria (Fig. 4). Their origin is unclear, since they branch off at the base of the tree (Fig. 3). Among those, *Herpetosiphon aurantiacus* harbors also an animal-type amylase. Its relative *Chloroflexus aurantiacus*, which has also a clearly orthologous animal-type amylase, lacks the Dictyo-type one.

The second scenario is that the animal α-amylases did not appear of their own by internal dramatic evolution, but rather from elsewhere, putatively through horizontal transfer. The main possibility is opposite to our earlier scenario [11]. Actually, animal-type amylase could have come from some Bacteria which have such enzymes. Our present data lend some support to this hypothesis, because the absence of animal amylase in non-Bilaterian suggests an appearance of this type in Metazoa much more recent than previously thought. The most probable donor candidate could be a member of Alteromonadaceae. The reason is that α-amylases of *Pseudoalteromonas* species and the related *Saccharophagus degradans*, have a C-terminal domain (AHA), very different from the CBMs. In *P. haloplanktis*, it serves as an export signal trough the periplasmic wall [49]. To our knowledge, this domain has not been found elsewhere in Bacteria. In contrast, it is present in many animal α-amylases ([11] and JLDL, unpublished data) and appears to be ancestral in Bilateria. It could be considered as a marker of the origin of animal amylases. If this hypothesis is correct, one would have to stress that the distribution of the animal-type amylases in Bacteria is though scarce and scattered, as if it had been an innovation of a small bacterial group, followed by episodic and rare horizontal transfers among bacteria (Fig. 4B). On the other hand, if ancestral in Eubacteria, it would have been lost or rearranged beyond recognition in a majority of lineages.

Because of the current paucity of data on non-Bilaterian metazoa, we cannot rule out a last possibility: an animal-type amylase coexisted with the ancestral Dictyo-type amylase in non-Bilateria animals, but has been lost in the species sampled here. Its presence in animal genomes would thus be older than suggested here. The ancestor of Bilateria would therefore have retained the animal-type and lost the Dictyo-type amylase. Further data from other sponges, jellyfish and sea anemone species are needed to confirm or reject this hypothesis.

Acknowledgments: We thank all people who have released their sequence data before publication, two anonymous reviewers for excellent comments, and Cushla Metcalfe for improving English.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2007.07.019.

References

[1] Henrissat, B., Deleury, E. and Coutinho, P.M. (2002) Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? Trends in Genetics 18, 437–440.
[2] Daïnou, O., Cariou, M.L., David, J.R. and Hickey, D. (1987) Amylase gene duplication: an ancestral trait in the *Drosophila melanogaster* species subgroup. Heredity 59, 245–251.
[3] Da Lage, J.L., Lemeunier, F., Cariou, M.L. and David, J.R. (1992) Multiple amylase genes in *Drosophila ananassae* and related species. Genetical Research Cambridge 59, 85–92.
[4] Laulier, M. (1988) Génétique et systématique évolutives du complexe d'espèces *Sphaeroma hooker*i Leach, *Sphaeroma levii* Argano et *Sphaeroma rugicauda* Leach (Crustacés Isopodes Flabellifères). 1. Génétique formelle de onze locus enzymatiques. Genetics, Selection, Evolution 20, 63–74.
[5] Sutliff, T.D., Huang, N., Litts, J.C. and Rodriguez, R.L. (1991) Characterization of an α-amylase multigene cluster in rice. Plant Molecular Biology 16, 579–591.
[6] Stanley, D., Fitzgerald, A.M., Farnden, K.J.F. and MacRae, E.A. (2002) Characterisation of putative alpha-amylases from apple (*Malus domestica*) and *Arabidopsis thaliana*. Biologia Bratislava 57, 137–148.
[7] Nakajima, R., Imanaka, T. and Aiba, S. (1986) Comparison of amino acid sequences of eleven different alpha-amylases. Applied Microbiology and Biotechnology 23, 355–360.
[8] Janecek, S. (1994) Sequence similarities and evolutionary relationships of microbial, plant and animal alpha-amylases. European Journal of Biochemistry 224, 519–524.
[9] Janecek, S. (1997) Alpha-amylase family: molecular biology and evolution. Progress in Biophysics and Molecular Biology 67, 67–97.
[10] Pujadas, G. and Palau, J. (2001) Evolution of α-amylase: architectural features and key residues in the stabilization of the (β/α)8 scaffold. Molecular Biology and Evolution 18, 38–54.
[11] Da Lage, J.L., Feller, G. and Janecek, S. (2004) Horizontal gene transfer from Eukarya to Bacteria and domain shuffling: the α-amylase model. Cellular and Molecular Life Science 61, 97–109.
[12] Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3, 853–859.
[13] Henrissat, B. and Davies, G. (1997) Structural and sequence based classification of glycoside hydrolases. Current Opinion in Structural Biology 7, 637–644.
[14] Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M. and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of a amylase related proteins. Protein Engineering, Design & Selection 19, 555–562.
[15] Janecek, S. (2002) How many conserved sequence regions are there in the alpha-amylase family? Biologia Bratislava 57, 29–41.
[16] D'Amico, S., Gerday, C. and Feller, G. (2000) Structural similarities and evolutionary relationships in chloride dependent α-amylases. Gene 253, 95–105.
[17] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402.
[18] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.
[19] Abascal, F., Zardoya, R. and Posada, D. (2005) ProtTest: selection of best fit models of protein evolution. Bioinformatics 21, 2104–2105.

[20] Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52, 696–704.
[21] Huelsenbeck, J.P. and Ronquist, F. (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.
[22] Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinformatics 5, 150–163.
[23] Furla, P., Allemand, D., Shick, J.M., Ferrier-Pagès, C., Richier, S., Plantivaux, A., Merle, P.L. and Tambutté, S. (2005) The symbiotic Anthozoan: a physiological chimera between alga and animal. Integrative and Comparative Biology 45, 595–604.
[24] Reshev, L., Koren, O., Loya, Y., Zilber Rosenberg, I. and Rosenberg, E. (2006) The coral probiotic hypothesis. Environmental Microbiology 8, 2068–2073.
[25] Snell, E., Furlong, R. and Holland, P. (2001) Hsp70 sequences indicate that choanoflagellates are closely related to animals. Current Biology 11, 967–970.
[26] King, N. and Carroll, S.B. (2001) A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. Proceedings of the National Academy of Sciences of the USA 98, 15032–15037.
[27] Long, C.M., Virolle, M.J., Chang, S.Y., Chang, S. and Bibb, M.J. (1987) Alpha-amylase gene of *Streptomyces limosus*: nucleotide sequence, expression motifs, and amino acid sequence homology to mammalian and invertebrate alpha-amylases. Journal of Bacteriology 169, 5745–5754.
[28] Janecek, S., Svensson, B. and MacGregor, E.A. (2003) Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch binding domain. European Journal of Biochemistry 270, 635–645.
[29] Rodriguez Sanoja, R., Oviedo, N. and Sanchez, S. (2005) Microbial starch binding domain. Current Opinion in Microbiology 8, 260–267.
[30] Bapteste, E. et al. (2002) The analysis of 100 gene supports the grouping of three highly divergent amoeba: *Dictyostelium*, *Entamoeba*, and *Mastigameoba*. Proceedings of the National Academy of Sciences of the USA 99, 1414–1419.
[31] Song, J., Xu, Q., Olsen, R., Loomis, W.F., Shaulsky, G., Kuspa, A. and Sucgang, R. (2005) Comparing the Dictyostelium and Entamoeba genomes reveals an ancient split in the Conosa lineage. PLoS Computational Biology, doi:10.1371/journal.pcbi.0010071.
[32] Fahrni, J.F., Bolivar, I., Berney, C., Nassonova, E., Smirnov, A. and Pawlowski, J. (2003) Phylogeny of lobose amoebae based on actin and small subunit ribosomal RNA genes. Molecular Biology and Evolution 20, 1881–1886.
[33] Feller, G., Lonhienne, T., Deroanne, C., Libioulle, C., Van Beeumen, J. and Gerday, C. (1992) Purification, characterization and nucleotide sequence of the thermolabile alpha-amylase from the antarctic psychrotroph *Alteromonas haloplanktis* A23. Journal of Biological Chemistry 267, 5217–5221.
[34] Sumitani, J.I., Nagae, H., Kawaguchi, T. and Arai, M. (1998) *Bacillus* animal type α-amylase: cloning and sequencing of the gene, and comparison of the deduced amino acid sequence with that of other amylases. Journal of Fermentation and Bioengineering 85, 428–432.
[35] Janecek, S., Lévêque, E., Belarbi, A. and Haye, B. (1999) Close evolutionary relatedness of α-amylases from archaea and plants. Journal of Molecular Evolution 48, 421–426.
[36] Simpson, A.G.B., Inagaki, Y. and Roger, A.J. (2006) Comprehensive multigene phylogenies of Excavates protists reveal the evolutionary positions of "primitive" Eukaryotes. Molecular Biology and Evolution 23, 615–625.
[37] Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J. and Gray, M.W. (2005) The tree of eukaryotes. Trends in Ecology and Evolution 20, 670–676.
[38] Da Lage, J.-L., Cariou, M.-L. and David, J.R. (1989) Geographical polymorphism of amylase in *Drosophila ananassae* and its relatives. Heredity 63, 67–72.
[39] De Jong, G., Hoorn, A.J.W., Thörig, G.E.W. and Scharloo, W. (1972) Frequencies of Amylase variants in *Drosophila melanogaster*. Nature 238, 453–454.
[40] Hoorn, A.J.W. and Scharloo, W. (1978) The functional significance of amylase polymorphism in *Drosophila melanogaster*. V. The effect of food components on amylase and alpha glucosidase activity. Genetica 49, 181–187.
[41] Inomata, N. and Yamazaki, T. (2002) Nucleotide variation of the duplicated *Amylase* genes in *Drosophila kikkawai*. Molecular Biology and Evolution 19, 678–688.
[42] Strobl, S., Maskos, K., Betz, M., Wiegand, G., Huber, R., Gomis Ruth, F.X. and Glockshuber, R. (1998) Crystal structure of the yellow meal worm alpha-amylase at 1.64 Å resolution. Journal of Molecular Biology 278, 617–628.
[43] Prigent, S., Matoub, M., Rouland, C. and Cariou, M.-L. (1998) Metabolic evolution in alpha-amylases from *Drosophila virilis* and *D. repleta*, two species with different ecological niches. Comparative Biochemistry and Physiology 119B, 407–412.
[44] Karn, R.C. and Malacinski, G.M. (1978) The comparative biochemistry, physiology and genetics of animal alpha-amylases. Advances in Comparative Physiology and Biochemistry 7, 1–103.
[45] Ramasubbu, N., Ragunath, C. and Mishra, P.J. (2003) Probing the role of a mobile loop in substrate binding and enzyme activity of human salivary amylase. Journal of Molecular Biology 325, 1061–1076.
[46] Ramasubbu, N., Ragunath, C., Mishra, P.J., Thomas, L.M., Gyémànt, G. and Kandra, L. (2004) Human salivary α-amylase Trp58 situated at subsite −2 is critical for enzyme activity. European Journal of Biochemistry 271, 2517–2529.
[47] Qian, M., Ajandouz, E.H., Payan, F. and Nahoum, V. (2005) Molecular basis of the effects of chloride ion on the acid base catalyst in the mechanism of pancreatic alpha-amylase. Biochemistry 44, 3194–3201.
[48] Gueix, N. and Peitsch, M.C. (1997) SWISS MODEL and the Swiss PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.
[49] Feller, G., D'Amico, S., Benotmane, A.M., Joly, F., Van Beeumen, J. and Gerday, C. (1998) Characterization of the C-terminal propeptide involved in bacterial wall spanning of α-amylase from the psychrophile *Alteromonas haloplanktis*. Journal of Biological Chemistry 273, 12109–12115.
[50] Burki, F. and Pawlowski, J. (2006) Monophyly of Rhizaria and multigene phylogeny of unicellular Bikonts. Molecular Biology and Evolution 23, 1922–1930.
[51] Hampl, V., Horner, D.S., Dyal, P., Kulda, J., Flegr, J., Foster, P.G. and Embley, T.M. (2005) Inference of the phylogenetic position of the Oxymonads based on nine genes: support for Metamonada and Excavates. Molecular Biology and Evolution 22, 2508–2518.
[52] Baldauf, S.L., Roger, A.J., Wenk Siefert, I. and Doolittle, W.F. (2000) A kingdom level phylogeny of eukaryotes based on combined protein data. Science 290, 972–977.