Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals

Nathan Lo*, Hirofumi Watanabe and Masahiro Sugimura
National Institute of Agrobiological Sciences, 1–2 Owashi, Tsukuba 305–8634, Japan
*Author for correspondence (nathanlo@utenti.unimi.it).

Until recently, the textbook view of cellulose hydrolysis in animals was that gut-resident symbiotic organisms such as bacteria or unicellular eukaryotes are responsible for the cellulases produced. This view has been challenged by the characterization and sequencing of endogenous cellulase genes from some invertebrate animals, including plant-parasitic nematodes, arthropods and a mollusc. Most of these genes are completely unrelated in terms of sequence, and their evolutionary origins remain unclear. In the case of plant-parasitic nematodes, it has been suggested that their ancestor obtained a cellulase gene via horizontal gene transfer from a prokaryote, and similar suggestions have been made about a cellulase gene recently discovered in a sea squirt. To improve understanding about the evolution of animal cellulases, we searched for all known types of these enzymes in GenBank, and performed phylogenetic comparisons. Low phylogenetic resolution was found among most of the sequences examined, however, positional identity in the introns of cellulase genes from a termite, a sea squirt and an abalone provided compelling evidence that a similar gene was present in the last common ancestor of protostomes and deuterostomes. In a different enzyme family, cellulases from beetles and plant-parasitic nematodes were found to cluster together. This result questions the idea of lateral gene transfer into the ancestors of the latter, although statistical tests did not allow this possibility to be ruled out. Overall, our results suggest that at least one family of endogenous cellulases may be more widespread in animals than previously thought.

Keywords: animal cellulase; phylogeny; gene transfer; cellulose; digestion

1. INTRODUCTION

Cellulases are enzymes that hydrolyse the β-1,4 linkages of cellulose: a linear homopolymer of glucose and the most abundant organic compound on Earth. They are commonly divided into two functional types: β-1,4-endoglucanases (EGs), which cleave the substrate from within (EC 3.2.1.4), and cellobiohydrolases (CBHs), that attack cellulose from the ends of the molecule (EC 3.2.1.91). Interest in these enzymes has increased in recent decades due to their important role in the global carbon cycle, as well as their use in alternative fuel production and other industries (Tomme et al. 1995). Cellulases are found in 14 out of the 90 known glycosyl hydrolase families (GHFs), which are classified on the basis of sequence similarity and hydrophobic cluster analysis (Henrissat & Bairoch 1993) (see also http://afmb.cnrs-mrs.fr/~cazy/CAZY). The majority of the cellulase-containing GHFs are unrelated in terms of sequence and structure, indicating that the ability to hydrolyse cellulose has evolved independently on several occasions.

Owing to the widespread belief that animals cannot produce cellulases, research on these enzymes has traditionally focused on those produced by non-animal organisms such as bacteria and fungi. However, recent years have seen the isolation of genes encoding cellulases from some invertebrate animal taxa (reviewed in Watanabe & Tokuda 2001). Rather than being from the one GHF, these cellulases are from three structurally and presumably phylogenetically unrelated families: GHF5, GHF9 and GHF45. GHF5 genes are found in various plant-parasitic nematodes (Smart et al. 1998; Rosso et al. 1999), GHF9 genes are known from termites (Watanabe et al. 1998; Tokuda et al. 1999), cockroaches (Lo et al. 2000) and a crayfish (Byrne et al. 1999), and GHF45 genes are known from a beetle (Girard & Jouanin 1999) and a mollusc (Xu et al. 2001). Most recently, a GHF9 cellulase gene has been reported from the genome of the sea squirt Ciona intestinalis, a urochordate deuterostome (Dehal et al. 2002). The evolutionary origins of all these genes remain unclear, first because most studies have focused on their molecular biological characteristics and the biochemistry of their products, and second because they appear to be a rarity among animals. Indeed cellulase genes are not found in the completed genomes of human, mouse, rat, fly, mosquito or nematode. In the case of plant-parasitic nematodes, phylogenetic comparisons of amino acid sequences have led to the hypothesis that their ancestors obtained cellulase genes via horizontal gene transfer (HGT) from bacteria (Yan et al. 1998). Similarly, the cellulase genes from the sea squirt are considered to represent ‘functional innovations’ specific to the urochordate lineage, with the tentative assumption that they have been acquired horizontally (Dehal et al. 2002). Overall, strong evidence for horizontal transfer of cellulase genes from bacteria or fungi to an animal has yet to be provided.

To increase knowledge about the evolution of these genes in animals, we searched for sequences in GenBank, and then performed phylogenetic comparisons of all known animal cellulases with their homologues from non-animal organisms, including the first analyses, to our knowledge, of those from beetles, molluscs and sea squirts. A goal of the study was to determine which of the following two hypotheses can best explain the presence of cellulase genes in animals today: (i) that these genes were present in early metazoans, and have been inherited in a vertical fashion to some extant animals with loss in many others; and (ii) that cellulase genes in extant animals were obtained more recently via sporadic cases of HGT from non-animal organisms.

2. METHODS

GenBank searches using the terms ‘metazoan’ and ‘animal’ in combination with either ‘cellulase’, ‘EG’ or ‘CBH’ were performed to
3. RESULTS AND DISCUSSION

In addition to the animal cellulases mentioned above, two additional genes have recently been deposited in the database. The first encodes a GHF5 enzyme, PhEG, from the longicorn beetle Psacothea hilaris. This gene has recently been determined in our laboratory, and has been shown to be endogenous by the sequencing of genomic fragments (M. Sugimura, unpublished data). The second is a GHF9 gene from the abalone Haliotus discus, also confirmed to be endogenous (Suzuki et al. 2003). Animal cellulases are thus apparently confined to GHFs 5, 9 and 45, and phyllogenetic analyses were performed on members of these families only.

On the basis of sequence similarity, GHF5 members can be divided into at least five subfamilies (Tomme et al. 1995), and those from plant-parasitic nematodes have been shown to be members of subfamily 2 (Yan et al. 1998). Blasts of the longicorn beetle cellulase PhEG showed it was closely related to subfamily 2 sequences, and thus phyllogenetic analyses of GHF5 were performed on this subfamily (figure 1). Trees inferred from ML and MP analyses were similar, though less resolution was found in the latter.

Nematode and longicorn beetle sequences clustered together with more than 50% support in both ML and MP analyses (figure 1), suggesting that their last common ancestor harboured a GHF5 gene. These animal cellulases were found to be most closely related to sequences from the bacteria Pseudomonas fluorescens and Erwinia chrysanthemi, as well as from the rumen fungus Orpinomyces joyonii. A close relationship between these microbial cellulases and those from nematodes was also found during phyllogenetic analyses by Yan et al. (1998) and Davis et al. (2000). Two a priori hypotheses—(i) that nematode cellulases have been acquired by horizontal transfer and (ii) that beetle and nematode cellulases are
Figure 2. Alignment of the *Ciona intestinalis* putative cellulase (CiPCEL) with the mature form of *NtEG*, a β-1,4-EG from the termite *Nasutitermes takasagoensis*, and *HdEG66*, a β-1,4-EG from the abalone *Haliotis discus hannai*. Portions of the N-termini of CiPCEL and *HdEG66* are not shown; asterisks indicate stop codons at the 3′-ends of the genes that encode them. The CiPCEL sequence was translated from a cDNA deduced from two overlapping clones in the GenBank EST database (accession numbers: AL665505 and AL668167). The accession numbers for *NtEG* and *HdEG66* are AB013272 and AB092978, respectively. Diamonds, triangles and arrows represent phase 0, 1 and 2 introns, respectively. The presence of an intron in the sea-squirt, termite and abalone sequences is indicated by S, T and A, respectively. For the termite *NtEG* sequence, all introns are known (Tokuda *et al.* 1999; accession number AB019146) while for abalone, only a genomic fragment from the 3′-end of the gene is available (the beginning of which is indicated by a hash; accession number AB092979). The introns in CiPCEL were identified in a straightforward manner by blasting the CiPCEL cDNA against the *C. intestinalis* genome database (http://genome.jgi-psf.org/ciona4). The entire CiPCEL gene is found on scaffold 11. Three introns are in identical positions in the sea-squirt and termite genes, while a further three introns are in identical positions in the sea-squirt and abalone genes. Of these, one is present among all three organisms (S+T+A). Other arthropod (cockroach and crayfish) GHF9 gene genomic fragments show intron positions exactly the same as in *NtEG* (Byrne *et al.* 1999; Lo *et al.* 2000).

the result of vertical descent from a common ancestor (indicated by the topology in figure 1)—were compared using the Kishino–Hasegawa test. For the hypothesis (i), the bacterial sequences from *P. fluorescens* and *E. chrysanthemi* were placed as sister groups to those from nematodes (leaving other parts of the tree unchanged), and the likelihood score was calculated. This was found to be 10.19 units less than the score for the topology in figure 1. However, this difference was not found to be significant at the 5% level. It was therefore not possible to distinguish between the two aforementioned hypotheses using the sequences currently available. One notable difference between the cellulases of beetles and nematodes is the presence of a CBD in some nematode enzymes. These cellulose-binding modules, which are attached to the catalytic domain by a linker sequence, have been found to be most closely related to CBDs from bacterial rather than eukaryotic cellulases (Smant *et al.* 1998). While this is suggestive of transfer of a GHF5 bacterial cellulase gene into the last common ancestor of plant-parasitic nematodes, it is conceivable that these domains are present in other eukaryotic cellulases that have not yet been discovered. In the case that a GHF5 cellulase was present in the last common ancestor of beetles and nematodes, one possibility is that the CBD has been lost from beetle enzymes. In this regard, it is noteworthy that a number of nematode cellulases do not contain a CBD.

Phylogenetic analyses of GHF9—which in addition to those from animals contains bacterial, plant, slime mould and fungal sequences—resulted in trees with poorly resolved nodes (data not shown). Low phylogenetic resolution was also found among members of GHF45 (which
includes bacterial, fungal and animal sequences). This low resolution prompted us to look for other sequence characteristics that could provide evidence for or against the idea of vertical descent of cellulase genes from a common ancestor. Notably, in the GHF9 genes of termites and the sea squirt, for which full sequences are available, we found three introns to be in identical positions (figure 2). This fact, not reported to our knowledge until now, provides compelling evidence that a cellulase gene with the same intron positions was present in the last common ancestor of these organisms. Additionally, in a 3′ genomic fragment of the abalone GHF9 gene, all three intron positions are identical to that in the sea-squirt gene, while one is shared both with the sea-squirt and termite genes. Thus, representatives of the three main branches of bilaterian animals—deuterostomes (sea squirt), ecdysozoans (termite) and lophotrochozoans (abalone)—all contain GHF9 genes with introns in identical positions, suggesting that they have inherited this gene vertically over several hundred million years from a primitive metazoan ancestor. An examination of intron positions of GHF9 cellulase genes of slime mould, fungus and various plants revealed no similarity to those in animals (data not shown). A lack of available intron data in some GHF5 and GHF45 animal sequences precluded a comparison of these for shared positions.

Among non-animal lineages the highest level of cellulase diversity is present in eubacteria, which have representatives from 12 out of the 14 cellulase-containing GHFs. The second highest level of diversity is found within fungi (10 different cellulase types). Given that fungi are among the closest relatives of animals, we might expect the last common ancestors of these two lineages to have harboured genes encoding most of the 10 types of cellulase currently known from them. One could speculate that, even if half of the 10 types of cellulase in fungi were obtained by HGT after the radiation of these organisms, several cellulase types were present in the ancestors of fungi + animals. In this study, we could only find reliable evidence for the presence of GHF9 genes in primitive animals, and the evolutionary origins of those from GHF5 and GHF45 remain equivocal. The role of GHF9 genes in the last common ancestor of bilaterians is intriguing. The nature of this organism is the subject of continuing debate (reviewed in Erwin & Davidson 2002), though one possibility is that it was some kind of marine filter feeder. In this case, cellulases may have allowed the organism to derive energy from cellulose present in algae that it fed upon.

There has clearly been widespread loss of GHF9 cellulase genes in various bilaterian animal lineages whose ancestors evolved specialized diets such as carnivory, or in those that have become parasites. This is borne out by the lack of cellulase genes in the genomes of human, rat, mouse, fly, mosquito and nematode. While sea squirts have retained cellulase genes, other deuterostomes such as vertebrates appear to have lost them completely. As the earliest branching vertebrates consist mainly of carnivores, it could be expected that several as-yet unidentified GHF9 genes—and possibly those from other GHFs—are present in animal taxa whose ancestors maintained diets that included cellulose. It should no longer be assumed, as has often been the case in past studies, that cellulase activity present in an invertebrate animal is due entirely to microbes present in its digestive tract.

**Acknowledgements**

We thank Nick Vass Coleman and Gaku Tokuda for constructive criticisms on the manuscript. Some of the work was supported in part by the Promotion of Basic Research Activities for Innovative Biosciences fund from the Bio-Oriented Technology Research Advancement Institution and by a Pioneer Research Project Fund (no. PRPF-0022) from the Ministry of Agriculture, Forestry and Fisheries of Japan. N.L. is supported by a Japanese Science and Technology Agency Postdoctoral Fellowship.


