Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes

Andrew J. King[a], Simon M. Cragg[b], Yi Li[a], Jo Dymond[b], Matthew J. Guille[b], Dianna J. Bowles[a], Neil C. Bruce[a], Ian A. Graham[a], and Simon J. McQueen-Mason[a,1]

[a]Centre for Novel Agricultural Products, Department of Biology, University of York, PO Box 374, York YO10 5YW, UK; and [b]School of Biological Sciences, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth PO1 2DY, UK

Edited by Richard A. Dixon, The Samuel Roberts Noble Foundation, Ardmore, OK, and approved January 28, 2010 (received for review December 10, 2009)

The digestion of lignocellulose is attracting attention both in terms of basic research into its metabolism by microorganisms and animals, and also as a means of converting plant biomass into biofuels. Limnoriid wood borers are unusual because, unlike other wood-feeding animals, they do not rely on symbiotic microbes to help digest lignocellulose. The absence of microbes in the digestive tract suggests that limnoriid wood borers produce all the enzymes necessary for lignocellulose digestion themselves. In this study we report that analysis of ESTs from the digestive system of *Limnoria quadripunctata* reveals a transcriptome dominated by glycosyl hydrolase genes. Indeed, >20% of all ESTs represent genes encoding putative cellulases, including glycosyl hydrolase family 7 (GH7) cellobiohydrolases. These have not previously been reported in animal genomes, but are key digestive enzymes produced by wood-degrading fungi and symbiotic protists in termite guts. We propose that limnoriid GH7 genes are important for the efficient digestion of lignocellulose in the absence of gut microbes. Hemocyanin transcripts were highly abundant in the hepatopancreas transcriptome. Based on recent studies indicating that these proteins may function as phenoloxidases in isopods, we discuss a possible role for hemocyanins in lignin decomposition.

biofuels | cellulase | hemocyanin | phenoloxidase | wood degradation

Plant lignocellulose is one of the most abundant reserves of fixed carbon on the planet and is mostly composed of cell walls. Energy-rich polysaccharide polymers comprise up to 70% of plant cell walls and these can be broken down to produce sugars to fuel heterotrophic metabolism or for use in industrial processes such as fermentation. There has been renewed interest in organisms capable of digesting lignocellulose in recent years. Such organisms play an import role in the recycling of organic materials leading to the release of CO_2 as part of the natural carbon cycle. In addition, the quest for sustainable liquid biofuels produced from plant biomass has led to a surge of interest in enzyme systems for obtaining sugars for fermentation from this source (1, 2). Obtaining nutritional value from lignocellulose is challenging as not only is cellulose highly indigestible due to its crystalline nature but also, the polysaccharide network is sealed in a highly resistant coating of lignin. Lignin is an unusual biological polymer in that it is not the product of a polymerase, but rather is formed by relatively random free radical polymerization of monolignols and therefore lacks a repeat structure that can be targeted by lytic enzymes. The digestion of woody plant tissues requires at least partial disruption of the lignin before the polysaccharides can be accessed, and this is accomplished using oxidative enzymes by many fungal species (3). Efficient enzymatic digestion of cellulose typically requires three major types of enzyme (4). β-1,4-endoglucanses (EC 3.2.1.4) hydrolyse accessible intramolecular glycosidic bonds to produce free chain-ends. These chain-ends can be digested with β-1,4-exoglucanases or cellobiohydrolases (EC 3.2.1.91) that release soluble cellobiose. The cellobiose is then cleaved to form glucose by β-glucosidases (E.C. 3.2.1.21). An alternative mechanism exists in the anaerobe *Clostridium themocellum*, where cellodextrins produced by the β-1,4-endoglucanses are transported into the cell and then undergo phosphorolytic cleavage (5). Complete enzymatic digestion of lignocellulose also requires hemicellulases, which degrade matrix polysaccharides such as xylans, mannans, arabinans, and galactans (6).

The degradation of lignocellulose by free-living microorganisms plays an important role in carbon cycling and has been studied in a range of organisms including the anaerobic bacterium *Clostridium thermocellum* (7). *C. thermocellum* produces a multiprotein complex, the cellulosome, which contains a range of cellulases, hemicellulases, structural components, and carbohydrate binding domains (7). In aerobic conditions white rot fungi, such as *Phanaerochaete chyrosporium* (8), or brown rot fungi such as *Postia placenta* (9), are major degraders of woody biomass. *P. chyrosporium* secretes a range of free enzymes such as lignin peroxidases and glycosyl hydrolases to digest lignocellulose (8). In contrast, *P. placenta* possesses far fewer ligninases and glycosyl hydrolases, but instead appears to use free radicals generated by the Fenton reaction to break down woody substrates (9, 10).

The digestion of lignocellulose has been studied in a range of terrestrial animals. One common theme observed in these animals is that they require symbiotic relationships with microorganisms to provide at least some of the digestive capabilities required to survive on such a diet. Ruminants such as cattle contain a range of bacteria and fungi in their rumen that produce a diverse array of cellulolytic enzymes (11). Termite hindguts also contain diverse microbial communities including bacteria (12) and protists (13). The complexity of these gut microbial communities, and the great diversity of enzymes that they produce was highlighted in a recent metagenomic study of a termite hindgut (12). Although the digestion of lignocellulose by terrestrial animals has been widely studied, the potential of marine organisms in the digestion of lignocellulose has received less attention.

Large quantities of lignocellulose enter marine ecosystems particularly from river estuaries and mangrove forests, and provide an important food source for a range of organisms: Evidence of marine wood-degrading communities extends back to the early Jurassic (14). This wood provides food for benthic invertebrates, in particular, two families of specialist wood borers, which are important pests for marine construction using timber. These are teredinid bivalves, known as shipworms (15), and isopod

Author contributions: A.J.K., S.M.C., J.D., M.J.G., D.J.B., N.C.B., I.A.G., and S.J.M.-M. designed research; A.J.K., Y.L., J.D., and M.J.G. performed research; A.J.K., Y.L., and S.J.M.-M. analyzed data; and A.J.K., S.M.C., and S.J.M.-M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The *L. quadripunctata* glycosyl hydrolase sequences reported in this paper have been deposited in the GenBank database [accession numbers GU066826 (GH5A), GU066827 (GH5C), GU066828 (GH5E), FJ940756 (GH7A), FJ940757 (GH7B), FJ940758 (GH7C), FJ940759 (GH9A), FJ940760 (GH9B), and FJ940761 (GH9F)]. Hemocyanin sequences have been deposited as accessions GU166295 (Hc1), GU166296 (Hc2), GU166297 (Hc3), and GU166298 (Hc4).

[1]To whom correspondence should be addressed. E-mail: sjmm1@york.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0914228107/DCSupplemental.

crustaceans of the Limnoriidae, known as gribble (16). Both of these borers ingest wood, but they have contrasting approaches to lignocellulose digestion. Like termites, teredinid wood borers have mutualistic associations with cellulolytic microbes, thought to play a critical role in lignocellulose digestion (15, 17). Similarly, terrestrial isopods appear to require gut microbes to aid in lignocelluloses digestion (18). In contrast, the digestive tract of wood boring marine limnoriids appears to be effectively sterile with no resident microbes, although some are ingested with the wood (19, 20). This is a particularly striking situation because it both implies that *Limnoria* do not require microbial mutualists, and also that the gut environment effectively prevents microbial proliferation. Not only is the sterile digestive tract of these organisms unusual from the point of view of gut physiology, but it also offers a substantial biotechnological opportunity for identifying enzymes and reaction conditions for lignocellulose degradation.

To explore the digestive processes in *L. quadripunctata*, a wood borer from temperate waters, we produced an EST collection of gene transcripts from the hepatopancreas. In crustaceans, this organ is involved in a range of functions including the production of digestive enzymes, nutrient uptake, carbon-reserve storage, ion transport, osmoregulation, heavy metal sequestration, and production of the oxygen binding pigment hemocyanin (21).

Results and Discussion

Preparation of a *Limnoria* Hepatopancreas Unigene Set. *L. quadripunctata* has a long gut that is normally packed with wood particles (Fig. 1*A* and *B*). The hepatopancreas consists of two bilobed blind-ending sacs (Fig. 1*A*–*C*) that connect to the stomach region and are surrounded by a mesh of muscles, which are thought to contract to inject digestive enzymes into the gut (22). RNA was extracted from hepatopancreas lobes dissected from 30 specimens of *L. quadripunctata* collected from infested wood. cDNA synthesized from this material served as template for massively parallel DNA pyrosequencing, which yielded a total of 418,749 DNA sequences with an average length of 247 bp, giving a total of over 106 million bp of sequence. The sequences were assembled and aligned to reveal 12,306 contiguous sequences, leaving 51,683 singleton sequences that were not annotated. The collection of contiguous sequences was annotated by automated sequence alignments to the collection of nonredundant peptide sequences held in GenBank, using the BLASTx algorithm, which searches protein sequence databases with all six possible translations of a DNA sequence. Annotation revealed 11,974 nonribosomal contiguous sequences, of which 4,336 (36.2%) had e-values <0.001 in BLASTx searches, indicating close sequence similarity to known genes. A summary of the most common types of sequence is shown in Fig. 1*D*, and a list of the most abundantly represented sequences in Table 1.

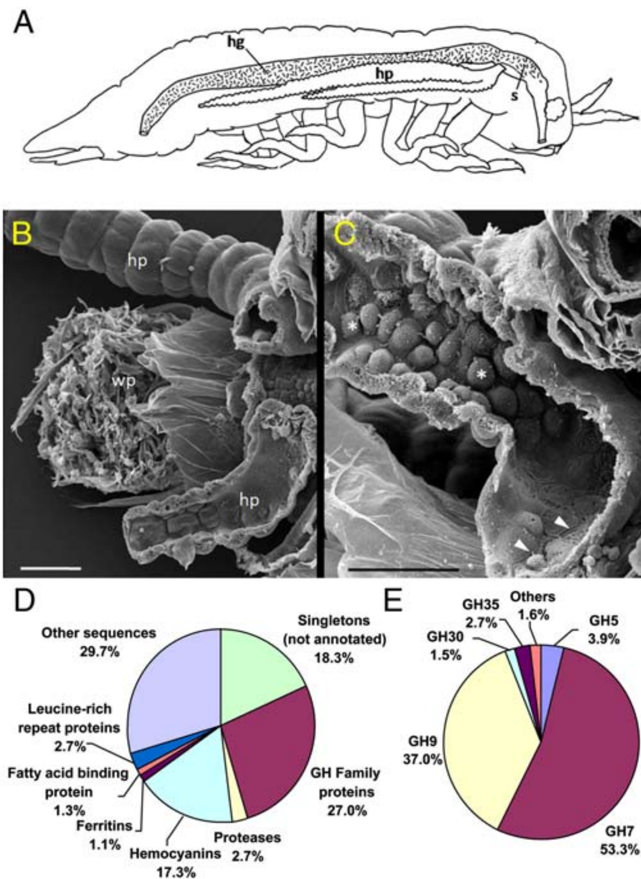

Fig. 1. Overview of the anatomy and summary of the digestive transcriptome of *L. quadripunctata*. (*A*) Lateral view of *Limnoria quadripunctata* showing the junction of one of the two pairs of hepatopancreas (hp) lobes with the ventral surface of the posterior region of the stomach(s) just anterior to the junction with the hindgut (hg). Food mass indicated by hatching. (*B*) SEM image of specimen of *L. quadripunctata* cut at the junction of the hg with the stomach, showing an intact lobe of the hp above, with contracted circular muscles, the mass of wood particles (wp) in the gut and a sectioned hp lobe below. The connection between the hp and the stomach was removed during preparation. Scale bar = 50 μm. (*C*) Lumen of hepatopancreas with swollen apical surfaces of individual cells (asterisks) and cell boundaries marked by differences in microvilli (arrowheads). Scale bar = 50 μm. (*D*) Diagrammatic overview of the most represented types of sequence amongst ESTs from the *L. quadripunctata* hp cDNA library. (*E*) distribution of ESTs corresponding to putative glycosyl hydrolase-encoding gene families GH7, GH9, GH5, GH35, GH30, and others (GH2, GH13, GH16, GH18, GH20, GH31, and GH38).

Glycosyl Hydrolases and Hemocyanins Dominate the *Limnoria* Hepatopancreas Transcriptome. The transcriptome is dominated by a relatively small number of sequence types. By far the most abundant sequences correspond to glycosyl hydrolases and hemocyanins respectively representing 27% and 17.3% of all ESTs. Proteases make up about 2.7% of the transcriptome. Homologues of a leucine-rich repeat protein represent 2.7% and ESTs with relatively low homology to a fatty acid binding protein account for 1.3%, whereas homologues of ferritin represent 1.1% of the transcriptome.

The EST collection includes sequences representing 12 recognizable glycosyl hydrolase families (Fig. 1*E*) following the CAZy nomenclature for carbohydrate active enzymes (23). The vast majority of the glycosyl hydrolase transcripts (94.2%) belong to three families, the GH7 (53.3%), GH9 (37.0%), and GH5 (3.9%), which have previously been shown to have a role in the digestion of lignocellulose. Although absent from most animal species, glycosyl hydrolases involved in lignocellulose digestions have previously been reported in a number of animals. Most well-characterized animal cellulases come from the GH9 family (24). We were able to distinguish at least six GH9 genes in the EST database, which together account for slightly less than 10% of the transcriptome. The *Limnoria* GH9 proteins show highest sequence similarity to known β-1,4-endoglucanases from arthropods such as termites (25) and crayfish (26) (Fig. 2*A*). EST databases from the freshwater amphipod, *Gammarus pulex* and the cladoceran water flea, *Daphnia magna* are also know to contain GH9 family proteins (24). The presence of GH9 family proteins in the EST library is therefore not unusual.

GH5 family proteins have also previously been reported in a number of animal species (24). The GH5 family contains a diverse range of glycosyl hydrolases from bacteria, fungi, plants, and animals. It includes proteins with both cellulase and hemicellulase type activities. Phylogenetic analysis (Fig. S1) reveals that the *Limnoria* GH5 proteins are most similar to a subfamily of GH5 proteins with β-1,4-mannase activity that are present in a

Table 1. BLASTx matches to the most abundantly represented ESTs

	Annotation	Organism	Function	E-value	GenBank	ESTs	%
1	Cellulase	*Pseudotrichonympha grassii*	Cellulose digestion (GH7 family)	1e-145	BAB69425	23,668	8.4
2	Putative glycosyl hydrolase family 7	Uncultured symbiotic protist of *Hodotermopsis sjoestedti*	Cellulose digestion (GH7 family)	2e-149	BAF57342	16,632	5.9
3	β-1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	6e-127	AAD38027	15,755	5.6
4	Hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0.0	CAI78901	13,323	4.7
5	Hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0.0	CAI78901	8,473	3.0
6	Hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0.0	CAI78901	7,245	2.6
7	Hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0.0	CAI78901	6,272	2.2
8	Hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0.0	CAI78901	5,001	1.8
9	Hemocyanin subunit 1	*Gammarus roeseli*	Oxygen transporter	0.0	CAI78901	3,766	1.3
10	β-1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	1e-130	AAD38027	3,549	1.3
11	Fatty acid-binding protein	*Bufo arenarum*	Unknown	5e-08	P83409	3,082	1.1
12	Hemocyanin	*Pacifastacus leniusculus*	Oxygen transporter	0.0	AAM81357	2,916	1.0
13	Ferritin peptide	*Fenneropenaeus chinensis*	Iron storage protein	1e-57	ABB05537	2,455	0.9
14	Hypothetical protein	*Branchiostoma floridae*	Unknown (Leucine-rich repeat)	1e-24	XP_002223116	2,330	0.8
15	Lysosomal β-galactosidase	*Canis lupus familiaris*	β-galactosidase (GH35 family)	2e-132	ABA43388	2,293	0.8
16	Chymotrypsin BII	*Litopenaeus vannamei*	Protease	9e-67	P36178	2,094	0.7
17	β-1,4-mannanase precursor	*Cryptopygus antarcticus*	Mannan digestion	1e-75	ABV68808	1,720	0.6
18	β-1,4-endoglucanase	*Cherax quadricarinatus*	Cellulose digestion (GH9 family)	7e-135	AAD38027	1,660	0.6
19	Trypsin	*Litopenaeus vannamei*	Protease	2e-65	CAA60129	1,546	0.6
20	Hypothetical protein	*Branchiostoma floridae*	Unknown (Leucine-rich repeat)	1e-24	XP_002223116	1,455	0.5

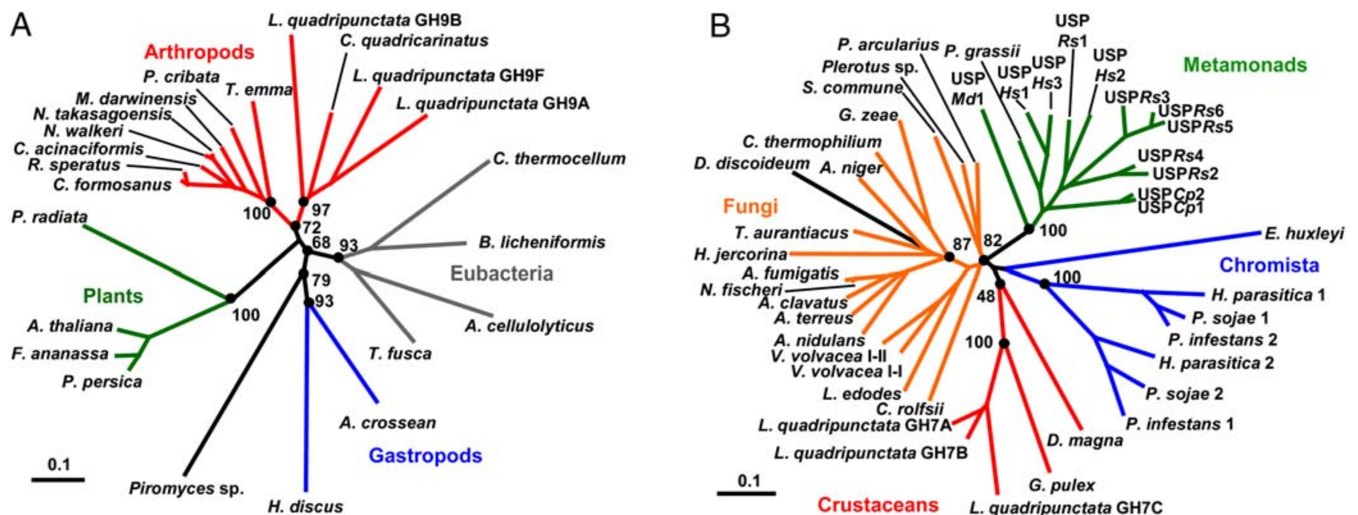

Fig. 2. Phylogenetic analysis of GH9 and GH7 family proteins from *Limnoria*. (*A*) Unrooted phylogenetic tree showing the relationship between GH9 family proteins of *L. quadripunctata*, other arthropods, gastropods, eubacteria, and plants. Arthropod sequences include *L. quadripunctata* GH9A (GenBank accession FJ940759), GH9B (FJ940760) and GH9F (FJ940761), *Cherax quadricarinatus* (AAD38027), *Teleogryllus emma* (ABV32557), *Panesthia cribrata* (AAF80584), *Mastotermes darwiniensis* (CAD54730), *Nasutitermes takasagoensis* (BAA33708), *Nasutitermes walkeri* (BAA33709), *Coptotermes acinaciformis* (AAK12339), *Reticulitermes speratus* (BAA31326), and *Coptotermes formosanus* (BAB40696). Gastropod sequences include *Ampullaria crossean* (ABD24280) and *Haliotis discus* (ABO26609). Eubacterial sequences include *Clostridium thermocellum* (ZP_03149988), *Bacillus licheniformis* (AAR29083), *Acidothermus cellulolyticus* (YP_873459), and *Thermobifida fusca* (YP_290232). Plant sequences include *Pinus radiata* (AAC12684), *Arabidopsis thaliana* (NP_192843), *Fragaria* x *ananassa* (AAC78298), and *Prunus persica* (CAI68019). A GH9 family cellulase from the chytridiomycote fungus *Piromyces* sp. is also included (AAM81967). (*B*) Unrooted phylogenic tree showing the relationship between GH7 family proteins of *Limnoria quadripunctata* and other crustaceans with those of ascomycete and basidiomycete fungi, chromista, and other protists. Ascomycete sequences include *Aspergillus clavatus* (GenBank Accession XP_001272622), *Aspergillus fumigatus* (XP_750600), *Aspergillus nidulans* (AAM54069), *Aspergillus niger* (AAF04492), *Aspergillus terreus* (XP_001212905), *Chaetomium thermophilum* (CAM98448), *Gibberella zeae* (AAR02398), *Hypocrea jercorina* Cel7A (P62694), *Neosartorya fischeri* (XP_001257823), and *Thermoascus aurantiacus* (CAM98447). Basidomycete sequences include *Corticium rolfsii* (JC7979), *Lentinula edodes* (AAK95563), *Pleurotus* sp. "Florida" (CAK18913), *Polyporus arcularius* (BAF80326), *Schizophyllum commune* (AAX55505), *Volvariella volvacea* cellobiohydrolase I-I (AAT64006), and cellobiohydrolase I-II (AAT64007). Chromista sequences include *Phytophthora infestans* CBH1 (AATU01001643, nt. 30669–32666) and CBH2 (AATU01002736, nt. 105664–107044), *Phytophthora sojae* CBH1 (AAQY01001052 nt. 13633–15024) and CBH2 (AAQY01000806 nt. 8540–10201), *Hyaloperonospora parasitica* CBH1 (ABWE01000089 nt. 41375–42991) and CBH2 (ABWE01000367 nt. 37087–38922), and *Emiliana huxleyi* (Fig. S3*C*). Crustacean sequences include *Limnoria quadripunctata* GH7A (FJ940756), GH7B (FJ940757), and GH7C (FJ940758); and *Daphnia magna* (Fig. S3*A*) and *Gammarus pulex* (Fig. S3*B*). Protist sequences include *Pseudotrichonympha grassii* (BAB69426), *Dictyostelium discoideum* (XP_647587) and a selection of uncultured symbiont protists isolated from the guts of the wood roach *Cryptocercus punctulatus* (USPCs1; BAF57469 and USPC*s*2; BAF57464), *Hodotermopsis sjoestedti*, (USP*Hs*1; BAF57344, USP*Hs*2; BAF57343 and USP*Hs*3; BAF57342), *Mastotermes darwiniensis* (USP*Md*; BAF57427), and *Reticulitermes speratus* (USP*Rs*1; BAF57302, USP*Rs*2; BAF57309, USP*Rs*3; BAF57308, USP*Rs*4; BAF57299, USP*Rs*5; BAF57312 and USP*Rs*6; BAF57297). All branches are drawn to scale as indicated by the scale bar (=0.1 substitutions per site). Bootstrap values (n = 1000) for the main branches are shown as percentages.

range of aquatic animals including molluscs (27, 28) and arthropods (29). It is therefore probable that the *Limnoria* ESTs also represent genes encoding β-1,4-mannanases, that could be involved in the digestion of hemicellulose.

The most abundantly represented glycosyl hydrolases in the EST collection belong to the GH7 family. We identified three distinct GH7 genes, which between them represented 14.4% of all the nonribosomal ESTs. Although cellulose digestion has been studied in many animals , no endogenous GH7 family glycosylhydrolase has previously been reported (30), so the presence of these sequences in the digestive transcriptome of *L. quadripunctata* is therefore surprising. GH7 enzymes are typically cellobiohydrolases, and are seen as critical enzymes for efficient cellulose digestion (31), but until now, have only been reported from fungi and from protozoan mutualists found in the digestive tract of termites (13, 32).

To determine the relationship between the limnorid GH7 proteins with other GH7s, we prepared a phylogenetic tree containing full-length peptide sequences obtained from the GenBank databases (see *Materials and Methods*) including a number of sequences obtained from the assembly of partial sequences obtained from EST and whole genome shotgun (WGS) sequences (Fig. 2). GH7 sequences were found in numerous ascomycete and basidiomycete fungal species, oomycetes, and several protist phyla, in particular, the Metamonada that include the oxymonads and parabasilids resident in the hindgut of many termite species (13). Although there have been no publications describing GH7 sequences from animals, we did find ESTs that encode such proteins from the freshwater amphipod *Gammarus pulex* and the recently released genome sequence of the water flea *Daphnia magna* both of which are crustacean relatives of *Limnoria* (Fig. 2, Table S1, and Figs. S2 and S3 *A* and *B*).

Endoglucanases working alone exhibit only slow release of soluble sugars from crystalline cellulose (31, 33). This is most probably because these enzymes can only cleave internal β-1, 4-glucans that protrude from amorphous regions in the crystalline cellulose. Even short β-1,4-oligoglucans are highly insoluble, and hence cleavage products remain associated with the cellulose. Cellobiohydrolases, such as those in GH7, processively cleave 2-glucose cellobiose units from the ends of glucan chains, but are also relatively ineffective in cellulose digestion if used alone. However, a combination of the two enzyme types is much more effective as the endoglucanase increases the number of exposed ends for the cellobiohydrolases to attack (31, 33). It is interesting to note that although termites possess their own β-1,4-endoglucanases (GH9) genes (25), proteomic analysis reveals that the most abundant proteins of the hindgut of *Reticulotermes* are GH7 cellobiohydrolases produced by protist endosymbionts (13).

The next most abundantly represented class of glycosylhydrolases in the EST database are the GH35 family and GH30 family. In eukaryotes, GH35s are typically β-galactosidases and may therefore be involved in degrading cell wall galactans or galactose side chains from galactomannans. Few eukaryotic GH30 proteins have been characterized, but this family does include glucosidases and xylosidase activities (34), raising the possibility that the *Limnoria* GH30s may be fulfilling either or both of these functions. Surprisingly, no obvious β-1-4-xylanase-like sequences were present in the transcriptome. A few examples of β-1-4-xylanases have been reported in animals, including the molluscs *Corbicula japonica* (35) and *Ampullaria crossean* (36). These belong to the GH10 family that was not detected in *Limnoria*. In prokaryotes, the GH5 family contains both β-1,4 mannanases and β-1-4 xylanases, and it is possible that the limnorid GH5 proteins may act on either or both of these substrates.

Validation of GH7 Sequences as Genuine Limnorid Genome Sequences. As no animals have previously been shown to possess GH7 genes in their genome, it was important to verify that the GH7 ESTs we describe here are actually expressed by *L. quadripunctata* rather than being produced by undetected gut microbes. A number of observations support the case that the GH7 sequences are of limnoriid origin. First, no other protist-like sequences are apparent in the EST annotations, and second, there is no sign of protist cells in either our own microscopic studies (Fig. 1*B* and *C*) or those published by other authors (19). Third, we carried out RNA in situ hybridization studies represented in Fig. 3. These studies show clear detection of transcripts for GH7 enzymes in the cytoplasm of large binucleate B cells of the hepatopancreas, confirming that the GH7 transcripts originate from the animal genome. The B cells are known to be secretory and show the variation in surface morphology evident in Fig. 1*C* during the secretory cycle (37).

It is possible that *Limnoria* acquired GH7 proteins through horizontal gene transfer, perhaps from association with a gut-resident organism. Horizontal gene transfer in metazoan species has now been demonstrated in a number of instances (38, 39). However, comparison of the GH7 protein sequences with their closest sequence homologues archived in GenBank show that they fall into a subgroup separate from either fungal or protist GH7s (Fig. 2*C*, protein alignments are presented as Fig. S2), and so it is unlikely that they originate from a recent gene transfer event.

The presence of GH7 sequences in other crustaceans is also interesting. In the case of *G. pulex*, the sequence similarity to the *Limnoria* GH7 proteins suggests that they are homologous to one another and are therefore likely to be encoded within the animal's genome rather than originating from endosymbionts. *G. pulex* is a detrivore whose diet includes lignocellulose. Although this species prefers partially decayed material, it is able to survive on diets of undecayed or "inactivated" leaf litter (40, 41). The hepatopancreas of *G. pulex* has been shown to contain endogenous cellulase (and phenoloxidase) activities that are not due to bacterial endosymbionts (41). It is not clear whether or not this species contains eukaryotic gut flora, but it is possible that the digestive system of *G. pulex* may share similarities with that of *L. quadripunctata*.

Could Hemocyanins Contribute to Lignin Degradation in *Limnoria*? In fungi, there are three major classes of enzymes involved in the breakdown of lignin; lignin peroxidases (LiP) and manganese peroxidases (MnP), which both contain Fe, and laccases, which are multicopper oxidases (3). At present, it is not known to what extent *L. quadripunctata* is able to degrade lignin. Although

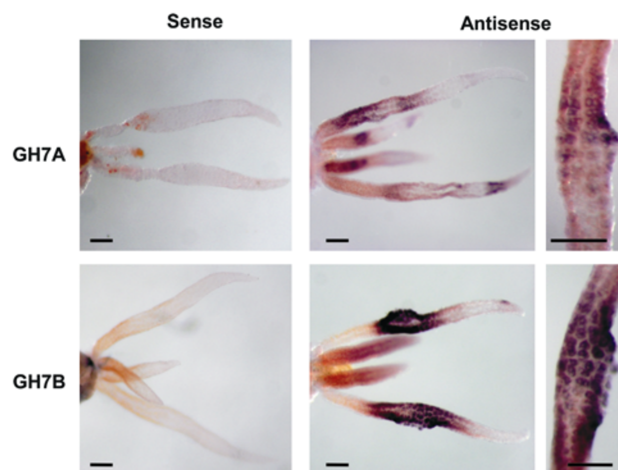

Fig. 3. Localization by RNA in situ hybridization of transcripts corresponding to *L. quadripunctata* GH7A and GH7B in formaldehyde-fixed hepatopancreas lobes. The sense probe acts as a control, whilst the antisense probe binds to the target transcript.

complete mineralization of lignin would not be necessary for a survival on a diet of wood, lignin presents a physical barrier to cellulases (41, 42). It therefore seems likely that this isopod will need to at least partly modify lignin to access the polysaccharides in wood. Furthermore, phenoloxidase activity has been observed in the digestive tracts of some crustaceans that feed at least partly on lignocellulose. The hepatopancreas of the amphipod *G. pulex*, which lacks bacterial endosymbionts, has been shown to contain phenoloxidase activity (41). Similarly, the coastal isopod *Ligia pallasii* is able to oxidize phenol compounds included in its diet, and this activity is not lost after antibiotic treatment, which suggests that the enzyme is produced by the isopod rather than gut microbes (43).

Phenoloxidases are required for a range of functions in crustaceans, including immune responses, and sclerotization and melanization of the exoskeleton (44), and in most cases, a prophenoloxidase located in the hemocytes is responsible for these activities. These prophenoloxidases are copper-containing enzymes that are closely related to the oxygen transporter hemocyanin. Interestingly, it has been demonstrated that the hemocyanins can be readily converted into phenoloxidases by conformational changes induced by limited proteolysis or chemical treatment (45). In vivo activation of both insect and crustacean hemocyanins has also been reported as an antimicrobial defense response (46, 47). There is mounting evidence to suggest the amphipods and isopods (peracarids) lack a *sensu strictu* prophenoloxidase, and instead it has been suggested that hemocyanins of these organisms, which diverged from other crustacean hemocyanins *ca.* 285 Mya (48) act as both oxygen carriers and phenoloxidases (44, 49, 50). We were not able to find ESTs for prophenoloxidases in the hepatopancreas transcriptome, nor for any other oxidases such as peroxidases or laccases that might be expected to fulfill a role in lignin degradation, but note that the hemocyanins ESTs are particularly abundant. Phylogenetic analysis reveals that these hemocyanins are closely related to those of other peracarids (Fig. 4 and Fig. S4). The abundance of these transcripts is not surprising, as one of the key functions of the hepatopancreas is hemocyanin production (51). However, considering that these proteins are thought to also be prophenoloxidases in the isopods, this does raise the possibility that they might be involved in lignin decomposition. In *Limnoria* and other peracarids, any hemocyanins present in the hepatopancreas could theoretically be activated into phenoloxidases by partial proteolysis or other factors. It is also interesting to speculate that the absence of microbes in the gut might be associated with the known antimicrobial effects of hemocyanin/phenoloxidases.

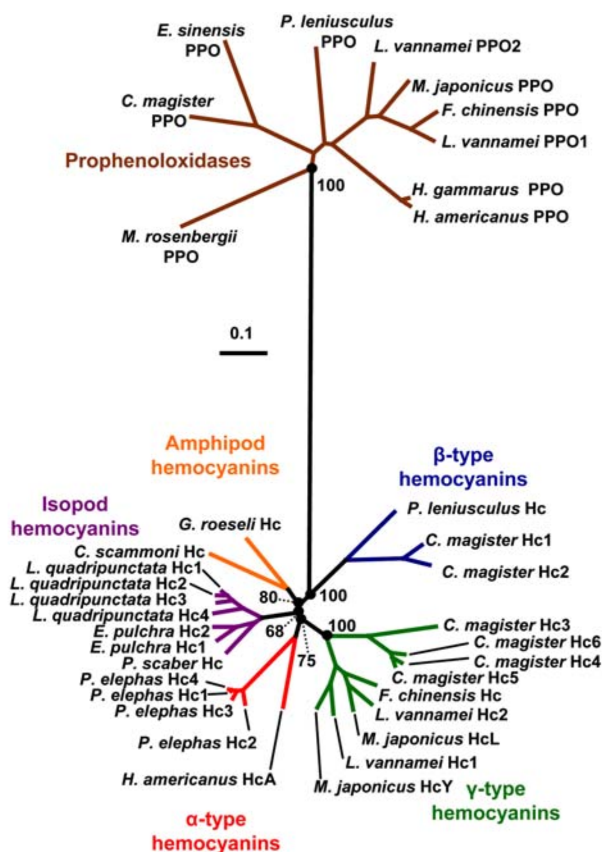

Fig. 4. Phylogenetic tree of crustacean hemocyanin and prophenoloxidase proteins. Eucarid hemocyanin (Hc) sequences include *Cancer magister* Hc1 (AAW57889), Hc2 (AAW57890), Hc3 (AAW57891), Hc4 (AAW57892), Hc5 (AAW57893) and Hc6 (AAA96966), *Pacifastacus leniusculus* (AAM81357), *Litopenaeus vannamei* Hc1 (CAA57880) and Hc2 (CAB85965), *Marsupenaeus japonicus* HcL (ABR14693) and HcY (ABR14694), *Fenneropenaeus chinensis* (ACM61982), *Palinurus elephas* Hc1 (CAC69243), Hc2 (CAC69244), Hc3 (CAC69245) and Hc4 (CAD56697), and *Homarus americanus* (CAB75960). Peracarid hemocyanins sequences include *L. quadripunctata* Hc1 (GU166295), Hc2 (GU166296), Hc3 (GU166297) and Hc4 (GU166298), *Eurydice pulchra* Hc1 (ACS44712) and Hc2 (ACS44713), *Porcellio scaber* (ACS44711), *Gammarus roeseli* (CAI78901), and *Cyamus scammoni* (ABB59715). Eucarid prophenoloxidase (PPO) sequences include *C. magister* PPO (ABB59713), *L. vannamei* PPO1 (ABX76968) and PPO2 (ABY81277), *Eriocheir sinensis* (ABS19633), *F. chinensis* (BAF98646), *Homarus gammarus* (CAE46724), *H. americanus* (AAT73697), *M. japonicus (BAB70485)*, *Macrobrachium rosenbergii* (ABA60740), and *P. leniusculus* (CAA58471). All branches are drawn to scale as indicated by the scale bar (=0.1 substitutions per site). Bootstrap values ($n = 1000$) for the main branches are shown as percentages.

Conclusion

This study has revealed a combination of glycosylhydrolase genes in *Limnoria* that seem likely to endow it with greater autonomous facility for lignocellulose digestion than animals such as termites. This may, in part, help to explain why these animals can survive on a diet of lignocellulose without the aid of gut microbes. The presence of GH7 family proteins in *Limnoria* is particularly striking, as they have not previously been reported as part of the endogenous cellulolytic systems of animals. Although a definitive study of the ability of *L. quadripunctata* to degrade lignin is required, the abundance of the hemocyanin transcripts, and the recently demonstrated ability of these proteins to function in vivo as phenoloxidases in isopods raises the possibility that these proteins could have a role in the degradation of lignin. We anticipate that further studies of the digestive system of these animals will provide some novel insights into lignocellulose degradation.

Materials and Methods

Diagram and Electron Microscopy. The animal outline in Fig. 1*A* was derived from an image of an animal in an artificial burrow, and the location of components of the digestive tract were observed from SEM images of sectioned animals. Specimens for SEM examination were fixed in sodium cacodylate-buffered glutaraldehyde, embedded in wax and sectioned. Wax was dissolved away from part-sectioned specimens and solvent traces were then removed by immersion in acetone then hexamethyldisilazane (HMDS). These specimens were dried by evaporation of the HMDS. Specimens were sputter-coated with gold and palladium. They were imaged in secondary electron mode at an acceleration voltage of 15 kV in a JEOL 6060LV SEM.

Production of a Transcriptome Database of the *L. quadripunctata* Hepatopancreas. Methods for the RNA isolation, cDNA synthesis, and EST database production are presented in *SI Text*.

In Situ Hybridization. Plasmid constructs containing nucleotide positions 52–370 (GH7A) or 51–370 (GH7B) were inserted into the BamHI/XhoI sites of pBluescript SK+ (Stratagene) using standard cloning techniques. Probes were prepared and the in situ hybridization performed on MEMFA fixed, dissected animals as described by Broadbent and Read (52). The method used was that for *Xenopus* but omitting the bleaching step.

Phylogenetic Analysis. Diverse members of the GH5, GH7, and GH9 protein families were obtained from a BLASTp search of the GenBank nonredundant peptide sequence databases using *Limnoria* peptide sequences as query sequences. For the GH7 family proteins, tBLASTn searches were also conducted against the GenBank nonredundant nucleotide, EST, and WGS collections. Contigs were assembled from the EST and WGS collections using Seqman II software (DNASTAR Inc.) and translated into proteins sequences using the ExPASy Translate tool (http://www.expasy.org/tools/dna.html). Protein sequence alignments were produced using ClustalX (53). Protein sequence alignments were then manually adjusted to minimize the effects of insertion/deletion events in the analysis. Phylogenetic and molecular evolutionary analyses were conducted using programs in the Phylip 3.6b package. Distance matrices were made with the protdist program using a Jones–Taylor–Thornton matrix and phylogenetic trees were then calculated from the matrices by the neighbor-joining algorithm. Bootstrap analyses used 1,000 replicates with the same protocol.

ACKNOWLEDGMENTS. We thank Graham Malyon for use of the image used to form the animal outline shown in Fig. 1*A*. This work was funded by Biotechnology and Biological Sciences Research Council Grant BB/G016178/1) and Leverhulme Foundation Grant F/00678/F, and by Amaethon Ltd.

1. Carroll A, Somerville C (2009) Cellulosic biofuels. *Annu Rev Plant Biol* 60:165–182.
2. Gomez LD, Steele-King CG, McQueen-Mason SJ (2008) Sustainable liquid biofuels from biomass: The writing's on the walls. *New Phytol* 178:473–485.
3. Wong D (2009) Structure and action mechanism of ligninolytic enzymes. *Appl Biochem Biotechnol* 157:174–209.
4. Percival Zhang YH, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol Adv* 24:452–481.
5. Zhang Y-HP, Lynd LR (2005) Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc Natl Acad Sci USA* 102:7321–7325.
6. Shallom D, Shoham Y (2003) Microbial hemicellulases. *Curr Opin Microbiol* 6:219–228.
7. Zverlov VV, Velikodvorskaya GA, Schwarz WH (2002) A newly described cellulosomal cellobiohydrolase, CelO, from *Clostridium thermocellum*: Investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology* 148:247–255.
8. Martinez D, et al. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* 22:695–700.
9. Martinez D, et al. (2009) Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci USA* 106:1954–1959.
10. Arantes V, Qian Y, Milagres AMF, Jellison J, Goodell B (2009) Effect of pH and oxalic acid on the reduction of $Fe3+$ by a biomimetic chelator and on $Fe3+$ desorption/adsorption onto wood: Implications for brown-rot decay. *Int Biodeter Biodeg* 63:478–483.
11. Duan CJ, et al. (2009) Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J Appl Microbiol* 107:245–256.
12. Warnecke F, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450:560–565.
13. Todaka N, et al. (2007) Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiol Ecol* 59:592–599.
14. Vahldiek B-W, Schweigert G (2007) Oldest record of wood-boring bivalves. *Neues Jahrb Geol P-A* 244:261–271.
15. Distel DL (2003) *Wood Deterioration and Preservation: Advances in Our Changing World*, eds B Goodell, DD Nicholas, and TP Schultz (Washington DC, American Chemical Society), pp 253–271.
16. Cragg S (2003) *Wood Deterioration and Preservation: Advances in Our Changing World*, eds B Goodell, DD Nicholas, and TP Schultz (American Chemical Society, Washington, DC), pp 272–288.
17. Yang JC, et al. (2009) The complete genome of *Teredinibacter turnerae* T7901: An intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS ONE* 4:e6085.
18. Wang Y, Brune A, Zimmer M (2007) Bacterial symbionts in the hepatopancreas of isopods: Diversity and environmental transmission. *FEMS Microbiol Ecol* 61:141–152.
19. Boyle PJ, Mitchell R (1978) Absence of microorganisms in crustacean digestive tracts. *Science* 200:1157–1159.
20. Daniel G, Nilsson T, Cragg S (1991) *Limnoria lignorum* ingest bacterial and fungal degraded wood. *Holz Roh Werkst* 49:488–490.
21. Brunet M, Arnaud J, Mazza J (1994) Gut structure and digestive cellular processes in marine Crustacea. *Oceanogr Mar Biol* 32:335–367.
22. Wägele J-W (1992) *Microscopic Anatomy of Invertebrates—Volume 9—Crustacea*, eds FW Harrison and AG Humes (Wiley–Liss Inc, New York), pp 529–618.
23. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* 37:D233–238.
24. Davison A, Blaxter M (2005) Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol* 22:1273–1284.
25. Watanabe H, Noda H, Tokuda G, Lo N (1998) A cellulase gene of termite origin. *Nature* 394:330–331.
26. Byrne KA, Lehnert SA, Johnson SE, Moore SS (1999) Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. *Gene* 239:317–324.
27. Ootsuka S, Saga N, Suzuki K-i, Inoue A, Ojima T (2006) Isolation and cloning of an endo-β-1,4-mannanase from Pacific abalone *Haliotis discus hannai*. *J Biotechnol* 125:269–280.
28. Larsson AM, et al. (2006) Three-dimensional crystal structure and enzymic characterization of β-mannanase Man5A from blue mussel *Mytilus edulis*. *J Mol Biol* 357:1500–1510.
29. Song JM, et al. (2008) Molecular cloning and characterization of a novel cold-active β-1,4-D-mannanase from the Antarctic springtail, *Cryptopygus antarcticus*. *Comp Biochem Phys B* 151:32–40.
30. Watanabe H, Tokuda G (2010) Cellulolytic systems in insects. *Annu Rev Entomol* 55:609–632.
31. Teeri TT (1997) Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends Biotechnol* 15:160–167.
32. Teeri TT, et al. (1998) *Trichoderma reesei* cellobiohydrolases: Why so efficient on crystalline cellulose?. *Biochem Soc Trans* 26:173–178.
33. Wilson DB (2008) Three microbial strategies for plant cell wall degradation. *Ann NY Acad Sci* 1125:289–297.
34. Brunner F, et al. (2002) A β-glucosidase/xylosidase from the phytopathogenic oomycete, *Phytophthora infestans*. *Phytochemistry* 59:689–696.
35. Sakamoto K, Toyohara H (2009) Putative endogenous xylanase from brackish-water clam *Corbicula japonica*. *Comp Biochem Phys B* 154:85–92.
36. Ding M, Teng Y, Yin Q, Zhao J, Zhao F (2008) The N-terminal cellulose-binding domain of EGXA increases thermal stability of xylanase and changes its specific activities on different substrates. *Acta Bioch Bioph Sin* 40:949–954.
37. Lešer V, et al. (2008) Epithelial thickness and lipid droplets in the hepatopancreas of *Porcellio scaber* (Crustacea: Isopoda) in different physiological conditions. *Zoology* 111:419–432.
38. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.
39. Rumpho ME, et al. (2008) Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica*. *Proc Natl Acad Sci USA* 105:17867–17871.
40. Willoughby LG, Sutcliffe DW (1976) Experiments on feeding and growth of the amphipod *Gammarus pulex* (L.) related to its distribution in the River Duddon. *Freshwater Biol* 6:577–586.
41. Zimmer M, Bartholmé S (2003) Bacterial endosymbionts in *Asellus aquaticus* (Isopoda) and *Gammarus pulex* (Amphipoda) and their contribution to digestion. *Limnol Oceanogr* 48:2208–2213.
42. Geib SM, et al. (2008) Lignin degradation in wood-feeding insects. *Proc Natl Acad Sci USA* 105:12932–12937.
43. Zimmer M, et al. (2002) Cellulose digestion and phenol oxidation in coastal isopods (Crustacea: Isopoda). *Mar Biol* 140:1207–1213.
44. Terwilliger NB, Ryan MC (2006) Functional and phylogenetic analyses of phenoloxidases from brachyuran (*Cancer magister*) and branchiopod (*Artemia franciscana*, *Triops longicaudatus*) crustaceans. *Biol Bull* 210:38–50.
45. Decker H, et al. (2007) Minireview: Recent progress in hemocyanin research. *Integr Comp Biol* 47:631–644.
46. Decker H, Jaenicke E (2004) Recent findings on phenoloxidase activity and anti microbial activity of hemocyanins. *Dev Comp Immunol* 28:673–687.
47. Jaenicke E, Decker H (2004) Functional changes in the family of type 3 copper proteins during evolution. *ChemBioChem* 5:163–169.
48. Hagner-Holler S, Kusche K, Hembach A, Burmester T (2005) Biochemical and molecular characterisation of hemocyanin from the amphipod *Gammarus roeseli*: Complex pattern of hemocyanin subunit evolution in Crustacea. *J Comp Physiol B* 175:445–452.
49. Pless DD, Aguilar MB, Falcón A, Lozano-Alvarez E, Heimer de la Cotera EP (2003) Latent phenoloxidase activity and N-terminal amino acid sequence of hemocyanin from *Bathynomus giganteus*, a primitive crustacean. *Arch Biochem Biophys* 409:402–410.
50. Jaenicke E, et al. (2009) Is activated hemocyanin instead of phenoloxidase involved in immune response in woodlice?. *Dev Comp Immunol* 33:1055–1063.
51. Kusche K, Burmester T (2001) Molecular cloning and evolution of lobster hemocyanin. *Biochem Bioph Res Co* 282:887–892.
52. Broadbent J, Read EM (1999) *Methods in Molecular Biology Volume 127: Xenopus and Zebrafish* (Humana Press, Totowa, New Jersey), pp 57–67.
53. Chenna R, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500.