The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer

Yannick Pauchet and David G. Heckel

Department of Entomology, Max Planck institute for Chemical Ecology, Hans-Knoell-Strasse 8, 07745 Jena, Germany

The primary plant cell wall comprises the most abundant polysaccharides on the Earth and represents a rich source of energy for organisms which have evolved the ability to digest them. Enzymes able to degrade plant cell wall polysaccharides are widely distributed in micro-organisms but are generally absent in animals, although their presence in insects, especially phytophagous beetles from the superfamilies Chrysomeloidea and Curculionoidea, has recently begun to be appreciated. The observed patchy distribution of endogenous enzymes encoding these enzymes in animals has raised questions about their evolutionary origins. Recent evidence suggests that endogenous plant cell wall degrading enzymes-encoding genes have been acquired by animals through a mechanism known as horizontal gene transfer (HGT). HGT describes how genetic material is moved by means other than vertical inheritance from a parent to an offspring. Here, we provide evidence that the mustard leaf beetle, *Phaedon cochlariae*, possesses in its genome genes encoding active xylanases from the glycoside hydrolase family 11 (GH11). We also provide evidence that these genes were originally acquired by *P. cochlariae* from a species of gammaproteobacteria through HGT. This represents the first example of the presence of genes from the GH11 family in animals.

1. Introduction

The primary cell wall, a heterogeneous mixture of polysaccharides and proteins, surrounds living plant cells and provide structural support as well as defence against pathogens. The primary cell wall is made of two polysaccharide networks: one contains cellulose microfibrils linked together by hemicellulose, and the other one is composed of pectins which hydrates and further cements the primary cell wall matrix [1]. Altogether these polysaccharides represent the most abundant bio-polymers on the Earth as well as a rich source of energy for organisms that have evolved the ability to degrade and assimilate them. Micro-organisms, particularly plant pathogenic fungi and bacteria, must degrade the primary cell wall before they can invade plant cells. To accomplish this they secrete a diverse and large arsenal of so-called plant cell wall degrading enzymes (PCWDEs). These enzymes were long thought to be absent in animals, until the first endogenous cellulase from a termite was characterized [2]. Since then, endogenous genes encoding many and diverse PCWDEs have been found in other animals such as plant-parasitic nematodes [3], springtails [4], mussels [5,6], bdelloid rotifers [7] and insects, particularly phytophagous beetles from the superfamilies Chrysomeloidea and Curculionoidea [8]. Interestingly, PCWDEs are absent from the genomes characterized to date of herbivorous insects and notorious crop pests from the order Lepidoptera (butterflies and moths; [9–11]). The so-called patchy distribution of endogenous PCWDE-encoding genes in animals has raised questions about their evolutionary origins. Whereas one gene family encoding cellulases from the glycoside hydrolase family 9 clearly has an ancient origin within animals [12], other gene families encoding PCWDEs were hypothesized to have been originally acquired by animals from micro-organisms through a mechanism known as horizontal gene transfer (HGT; [3,7]).
The process by which genetic material moves asexually across species boundaries, regardless of the evolutionary distance between them, is termed HGT. HGT is recognized as a major feature of genome evolution in prokaryotes [13], and its importance in the evolution of unicellular asexual eukaryotes is finally being acknowledged [7,14]. By contrast, evidence supporting HGT in animal genomes has often been rebutted, mainly due to key aspects of animal reproduction. In fact, for a foreign gene to have a chance to be propagated and fixed in a population, the HGT event has to target the germ line; and the separation of somatic and germ cells in animals represents a tremendous obstacle to HGT [15,16]. It is therefore not surprising that the vast majority of HGT events reported to date in animals have arisen from endosymbionts living predominantly in germ cells, such as the bacteria Wolbachia [15].

Nonetheless, although the mechanism remains elusive, the number of bacteria-to-animal HGT events arising from free-living bacteria, rather than endosymbionts, is currently increasing rapidly, together with the number of animal genomes and transcriptomes being sequenced. This is particularly true in two phylogenetically closely related phyla, nematodes and insects, in which physiologically and ecologically important functional HGTs have been recently described [17–21]. As an example, plant-parasitic nematodes have acquired genes encoding PCWDEs from several families; these genes, after further duplication events, have promoted their ability to parasitize plants [3]. Although we suggested in our previous work the possibility that genes encoding PCWDEs in phytophagous beetles may have arisen from HGT events [22] similar to plant-parasitic nematodes, only recently was the first strongly supported case of a functional and adaptive HGT of a mannanase gene in the coffee berry borer, Hypothenemus hampei, reported [23]. Larvae of the coffee berry borer feed exclusively inside coffee berries, which are primarily composed of polysaccharides including a high proportion of mannan. The acquisition of a mannanase gene through HGT may have been a key mechanism by which this species has adapted to its host plant and become a major pest of coffee worldwide.

Here, we functionally characterized two active xylanases from the glycoside hydrolase family 11 (GH11), a class of PCWDEs able to degrade xylan, the predominant constituent of the hemicellulose matrix of the plant cell wall and the second most abundant polysaccharide on the Earth, derived from the mustard leaf beetle, Phaedon cockleariae. These two GH11 genes were originally discovered by analysing the larval gut content proteome of this beetle species [24]. We also provide evidence that these genes were originally acquired by the insect through HGT, arising most probably from a species of gammaproteobacteria. After the initial HGT in the beetle genome, the two GH11 genes arose from a tandem gene duplication event. Xylanases from the GH11 family are widely distributed in micro-organisms but are generally absent in animals. Therefore, the presence of two endogenous genes encoding functional GH11 xylanases in a leaf beetle represents a unique example in animals.

2. Material and methods

(a) Leaf beetles

Phaedon cockleariae was collected on Brassicaceae close to the city of Bayreuth (Germany). Larvae and adults were kept as a continuous culture in the laboratory, at 20°C and on a cycle of 16 L:8 D, on leaves of Brassica rapa chinensis.

(b) Phylogenetic analyses

Protein sequences belonging to the GH11 family were collected from the CaZy database (http://www.cazy.org; [25]). Amino acid alignments were carried out using MUSCLE v. 3.7 on the Phylogeny.fr web platform (http://www.phylogeny.fr; [26]) and were inspected and corrected manually when needed. Phylogenetic relationships were estimated using three methods: maximum-likelihood, Bayesian inference and neighbour-joining. Phylogenetic analyses using both maximum-likelihood and neighbour-joining methods were conducted in MEGA5 [27]. For maximum-likelihood analysis, the best model of protein evolution was determined in MEGA5 using the ‘find best DNA/protein models’ tool. The best model was the ‘Whelan and Goldman’ (WAG) model, incorporating a discrete gamma distribution (shape parameter = 5) to model evolutionary rate differences among sites (+G) and a proportion of invariable sites (+I). For the neighbour-joining analysis, the Jones–Taylor–Thornton model incorporating a discrete gamma distribution (shape parameter = 5) to model evolutionary rate differences among sites (+G) was chosen. The robustness of both analyses was tested using 1000 bootstrap replicates. Bayesian-inferred analyses were carried out in MrBayes v. 3.1.2 which was set for the amino acid models to mix, thereby allowing model jumping between fixed-rate amino acid models to select the best one (WAG, in our case). Markov chain Monte Carlo runs were carried out for 1 000 000 generations, after which log likelihood values showed that equilibrium had been reached after the first 400 generations in all cases; those data were subsequently discarded from each run and considered as ‘burn in’. Two runs were conducted for the dataset showing agreement in topology and likelihood scores.

(c) Genomic library construction and screening

Genomic DNA isolated from P. cockleariae pupae was used for fosmid library construction performed with the CopyControl fosmid library production kit (Epiconcept Biotecnologies), following the manufacturer’s instructions. The pCC1FOS vectors were packaged with MaxPlax Lambda extracts (Epiconcept Biotecnologies). The phage particles were used for EPI300-T1 cell infection. Stocks of infected cells were sent to ImaGenes (Berlin, Germany) for plating, stock library production of each clone and duplicate colony spotting on nylon membranes. Genes-specific primers (see the electronic supplementary material, table S1) were designed to amplify by polymerase chain reaction (PCR) the whole open reading frame (ORF) of both gh11-1 and gh11-2 genes, using genomic DNA prepared from P. cockleariae pupae. These fragments were further used to screen the fosmid library. Nylon membranes were individually hybridized with probes specific to both gh11-1 and gh11-2 genes using the Amersham ECL direct nucleic acid labelling and detection system (GE Healthcare) following the manufacturer’s protocol for probe labelling, hybridization and detection. A positive clone was amplified for fosmid preparation followed by shearing on a HydroShear DNA shearing device (GeneMachines) and cloned into SmaI-digested pUC19 (Fermentas) for shotgun sequencing. To screen the fosmid clone for repetitive elements as well as transposable elements (TEs), we used the web interface of RepeatMasker and the protein-based RepeatMasking tool (http://www.repeatmasker.org/). The complete sequence of the positive fosmid clone was submitted to GenBank with accession number: KC521546.

(d) Heterologous expression in insect cells

gh11-1 and gh11-2 ORFs were amplified from cDNAs using gene-specific primers (see the electronic supplementary material, table
We used the amino acid sequences of both GH11-1 and -2 as queries for BLAST searches against the GenBank non-redundant protein database. The top hits were in both cases sequences derived from various bacteria with up to 69 per cent identity on the amino acid level. This result was not unexpected as putative xylanases from the GH11 family are mostly distributed in saprophytic and plant pathogenic fungi and bacteria, and are generally absent in plants and animals. A few sequences derived from archaea and protozoa were also found in public databases. The two GH11s derived from *P. cochleariae* are therefore so far unique to this species and to animals in general. In fact, no homologues of these GH11s could be identified in deep coverage larval gut transcriptomes of closely related leaf beetles of the subfamily Chrysomelinae, such as the Colorado potato beetle, *Leptinotarsa decemlineata*, or the green dock beetle, *Gastrophysa viridula* [22]. GH11s are also absent from the genome of the red flour beetle, *Tribolium castaneum*, as well as from the recently sequenced genome of the bark beetle *Dendroctonus ponderosae* [29]. Notably, an active xylanase was recently characterized from a curculionid beetle, the coffee berry borer, *H. hampei*, but it is part of a different glycoside hydrolase family (GH10 instead of GH11; [30]). This suggests two possibilities: either (i) GH11 genes were present in the last common ancestor of eukaryotes and died out in all lineages, except for fungi and species of the genus *Phaedon* or even *P. cochleariae* itself, or (ii) these genes were acquired horizontally by the beetle or one of its ancestors from either a fungal or a bacterial source. Although the first hypothesis seems unlikely because it would imply massive independent gene losses in many eukaryote lineages, we decided to test it by reconstructing the phylogenetic relationships of *P. cochleariae* GH11-1 and -2 with those derived from fungi, bacteria, protozoa and archaea.

Phylogenetic analyses of the GH11 catalytic domains were performed using three different methods that clustered GH11-1 and -2 together with sequences derived from species of gamma-proteobacteria, more specifically from the genera *Cellulobrio* and *Teredinibacter* (figure 1). Most *Cellulobrio* species are saprophytic soil bacteria known for their ability to degrade plant cell wall polysaccharides [31]. *Teredinibacter turnerae* is an intracellular endosymbiont of wood-boring marine bivalves of the family Teredinidae (shipworms) known to degrade cellulose and other plant cell wall components [32]. This clustering is well supported by the three methods we used with very high posterior branch probabilities and bootstrap values, strongly supporting the hypothesis that HGT occurred between an ancestor of the beetle, or *P. cochleariae* itself and a species of gammaproteobacteria. Notably, the bacteria-derived proteins often harbour additional features compared with the beetle-derived ones, such as carbohydrate binding domains or other catalytic domains (GH5 or polysaccharide deacetylase), attached to the GH11 catalytic domains (see the electronic supplementary material, figure S1). Altogether, this implies that the original horizontally transferred gene contained only a GH11 catalytic domain or that the putative-associated domains were acquired by the beetle or one of its ancestors from either a fungal or a bacterial source. Although the first hypothesis seems unlikely because it would imply massive independent gene losses in many eukaryote lineages, we decided to test it by reconstructing the phylogenetic relationships of *P. cochleariae* GH11-1 and -2 with those derived from fungi, bacteria, protozoa and archaea.

3. Results and discussion

(a) PcoGH11s are closely related to xylanases from proteobacteria

We recently identified two putative xylanases and their corresponding transcripts by assaying the larval gut content of *P. cochleariae* for activity against several plant cell wall polysaccharides combined with a proteomic analysis and transcriptome sequencing [24]. These proteins are members of the glycoside hydrolase family 11 (GH11), according to the carbohydrate-active enzyme nomenclature [25], and we named them GH11-1 and GH11-2; one of the two corresponding transcripts encoding GH11-1 corresponded to one previously identified from a *P. cochleariae* cDNA library [28]. Both gh11-1 and gh11-2 transcripts harbour typical features of eukaryotic genes such as 5′- and 3′-untranslated regions as well as a poly(A) tail. Both predicted proteins share 79.7 per cent amino acid identity, possess a 17-amino acid N-terminal signal peptide, and harbour two catalytic glutamate residues characteristic of GH11 enzymes, suggesting that both corresponding proteins are active enzymes. GH11-1 possesses two predicted N-glycosylation sites, whereas GH11-2 has only one. Additionally, RNA-SEQ and quantitative PCR experiments indicated that both gh11 genes were highly expressed in the insect gut compared with the rest of the body, in larvae as well as in adults, pointing to a digestive function [24].

S1) and were cloned into the pMIB/V5-His vector B, in frame with a V5-(His)₆ epitope at the carboxyl-terminus. Positive constructs were then transfected in High Five cells (Invitrogen) using Insect GeneJuice (Novagen) as a transfection reagent. Stable cell lines expressing both GH11-1 and -2 were further selected by treating transfected cells with 80 µg ml⁻¹ blasticidin (Invitrogen). Expression of both constructs was verified by western blot using the anti-V5-HRP antibody (Invitrogen). N-glycosylation of heterogeneous proteins was assessed by treating culture medium with peptidase-n-glycosidase F (PNGase F, New England Biolabs) according to the manufacturer’s protocol, followed by western blot using the Anti-V5-HRP antibody.

(e) Xylanase activity assays

Culture medium from stable cell lines expressing GH11-1 and -2 was collected, concentrated five times using 20 ml concentrators (Pierce) and desalted using Zeba desalt spin columns (Pierce). Samples (2.5 µl) were prepared for zymogram by diluting them in Laemmli sample buffer without reducing agent before being run on a 12.5 per cent SDS–PAGE gel containing 0.1 per cent beechwood xylan. Electrophoresis was carried out at 4 °C using pre-chilled running buffer. Gels were then washed three times in a 2.5 per cent Triton X-100 solution for 15 min at 4 °C. Gel slices were equilibrated in reaction buffers of various pH values (from pH 2.6 to 7.6: McIlvaine buffer; pH 8.5: Tris–HCl buffer; pH 9.5: sodium carbonate buffer) for 16 h at 4 °C, followed by a 1 h incubation at 30 °C. Activity was revealed by transferring gel slices in a 0.1 per cent Congo Red solution for 2 h at room temperature and then destained with 1 M NaCl until pale activity bands appeared on a dark red background. For diffusion assays, culture medium samples were dialyzed overnight against McIlvaine buffer pH 4.6. Ten microliters were deposited in 2 mm holes made in a 1 per cent agarose Petri dish containing 0.1 per cent beechwood xylan and 50 mM McIlvaine buffer pH 4.6. Activity was revealed with Congo Red as described above.
HGT event between two species of protozoa, *Polyplastron multivesiculatum* and *Epidinium ecaudatum*, and a bacterium of the class Firmicutes. This finding, at least for *P. multivesiculatum*, has already been reported [34].

(b) *PcoGH11*s are encoded by the beetle genome

It is now well established that symbiotic protists as well as microorganisms that are part of the gut microbiota can contribute greatly to the degradation of plant biomass in some insects, especially termites and wood-feeding cockroaches, by producing various hydrolyzing enzymes targeting plant cell wall polysaccharides [8,35]. In order to test whether both *gh11* genes are encoded by the beetle genome itself and to exclude the possibility of contamination of our beetle transcriptome by transcripts of putative symbiotic origin, we screened a *P. cochleariae* fosmid library with specific probes designed for both *gh11-1* and *gh11-2* genes. Both probes hit a single 37.7 kb clone (23G16) which was retrieved and further sequenced to completion. Neither gene has an intron, and both were found in tandem, with 3′ ends separated by only 2109 bp of non-coding DNA from stop codon to stop codon (figure 2). Such placement indicates that both genes originated from a tandem gene duplication event. Owing to the high sequence conservation between *GH11-1* and *GH11-2* and the proximity of both genes, it is reasonable to believe that this tandem gene duplication event happened relatively recently, most probably after the acquisition by the beetle genome through HGT. No flanking genes could be identified on the rest of the fosmid clone, excluding any possible comparison with the *Tribolium* genome. Owing to the very low gene density observed on the fosmid, a contamination of the fosmid library by bacterial-derived DNA can be readily excluded. A screen using the REPEATMASKER software, which analyses the presence of interspersed repeats and low complexity DNA sequences, revealed that this fosmid clone is almost entirely covered with both class I (retrotransposons) and class II (DNA transposons) TEs (see the electronic supplementary material, table S2). More precisely, seven TEs, six from class I and a single one from class II, could be clearly identified. Homology-based BLAST searches against a non-redundant protein database at NCBI indicated that five of the seven TEs are closely related to insect-derived ones, whereas...
the last two show top hits for TEs of the freshwater snail, *Biomphalaria glabrata*. One long terminal repeat (LTR) retrotransposon shares homology to insect-derived class I TEs of the superfamily Bel-Pao. This superfamily of LTR retrotransposons is restricted to species part of the Metazoa [36], therefore excluding the possibility that this fosmid clone could be a part of a piece of DNA from putative symbiotic organisms (protozoa or fungi) of *P. cochleariae*. Altogether, this strongly supports the fact that the two *gh11* genes are indeed part of the beetle genome. This situation is similar to what has been observed for a gene encoding an active mannanase (*HhMAN1*) acquired via HGT in the coffee berry borer, another coleopteran insect.

The *HhMAN1* gene is flanked by TEs of the Tc1/mariner and hAT superfamilies [23]. TEs are often associated with genome recombination, mobilization, and reorganization. HGTs of TEs between several eukaryotes have also been reported [37]. Taking this into account and noting the association between *HhMAN1* and TEs, Acuna et al. [23] proposed that one component of the remaining elusive mechanism of bacteria-to-animal HGTs could be the transposon-mediated incorporation of foreign DNA. This would be conceivable only if the TEs associated with *HhMAN1*—or, in our case *gh11*-1 and *gh11*-2—have some similarities with prokaryotic-derived ones [38]. However, this did not turn out to be true in either case: the TEs surrounding *HhMAN1* and the two *gh11* genes show similarities to other insect TEs rather than to prokaryotic-derived ones. But in the case of the *P. cochleariae* *gh11* genes, the repeated sequences typically found at the beginning and at the end of TEs may have promoted the tandem gene duplication event by facilitating an unequal crossing over [39].

(c) *PcoGH11* genes encode active xylanases

We previously demonstrated that xylanase activity was present in *P. cochleariae* larval gut contents, which led to the initial identification of *GH11-1* and *-2* [24]. Both proteins were successfully expressed in insect cells and were found to be secreted in the culture medium (figure 3a) and at comparable levels (see the electronic supplementary material, figure S2). PNGase F treatment also demonstrated that both heterologously expressed proteins are glycosylated (figure 3b). Zymograms on semi-native gels, using beech-wood xylan as a substrate, revealed that both enzymes possess the ability to degrade this substrate with a similar pH optimum around 4.5–5, although the observed activity.

![Figure 2](http://rspb.royalsocietypublishing.org/lookup/suppl/doi:10.1098/rspb.2013.1021/-/DC1/Fig2.jpg)

**Figure 2.** *Pcogh11*-1 and -2 are integrated in the *P. cochleariae* genome. Genome organization of both GH11 genes was obtained by sequencing a positive fosmid clone after screening a *P. cochleariae* fosmid library with specific probes designed for both genes. Both genes are found facing each other and surrounded by several types of transposable element.

![Figure 3](http://rspb.royalsocietypublishing.org/lookup/suppl/doi:10.1098/rspb.2013.1021/-/DC1/Fig3.jpg)

**Figure 3.** Heterologous expression of *GH11*-1 and -2 in insect cells. (a) Both *GH11*-1 and -2 were stably expressed in High Five cells. Cell lysates (L) and corresponding culture media (M) were tested for expression by western blot using an anti-V5-HRP antibody. Both heterologously expressed proteins were tested by treating corresponding culture media with PNGase F (see §2 for details) followed by western blot using an anti-V5-HRP antibody. The size shift observed between untreated (C) and PNGase F-treated (T) culture media indicates that both proteins are glycosylated.
of GH11-1 is far below that of GH11-2 (figure 4a). However, diffusion assays on agarose plates containing beechwood xylan as a substrate with non-denatured enzymes indicate that GH11-1 and -2 have a similar level of activity (figure 4b). The discrepancy in enzymatic activity observed on semi-native gels may be due to a poor renaturation of GH11-1 in the condition of the zymogram assay.

Altogether, our data indicate (i) that the xylanase activity detected in *P. cochleariae* gut contents [24] represents the contribution of both GH11-1 and -2 and (ii) that both GH11-1 and -2 kept their ancestral function after the tandem gene duplication event. In addition, both gh11 genes have similar expression pattern indicating that, after fixation, neither duplicate evolved by neo-functionalization or by sub-functionalization. For some genes, duplication provides an immediate selective advantage by increasing the expression level and thus the amount of protein (enzyme) produced, which is most probably the case here for the two *P. cochleariae* gh11s [39]. Another possibility would be that one of the two gh11 copies evolved to produce an enzyme which would be less sensitive to the putative xylanase inhibitors produced by the host plant, if produced at all, as a defence mechanism. There are three described categories of plant-derived proteinaceous inhibitors of xylanases-targeting enzymes from the GH10 and GH11 families: the TAXI-type, the XII-type and the thaumatin-like type (TL-XI-type) [40]. These three categories of xylanase inhibitors are part of relatively large gene families and have, so far, only been described in species of cereals [40]. We screened the published *Brassica rapa* genome [41], a host plant of *P. cochleariae*, for genes putatively encoding any of these three categories of xylanase inhibitors, by performing BLAST searches using known xylanase inhibitor sequences as queries. This screening did not provide any strong evidence for the presence of xylanase inhibitors in this plant species.

### 4. Conclusions

We believe that natural selection plays a crucial role in the maintenance of genes acquired by HGT in the recipient genome. Some bacteria-to-animal HGTs are clearly adaptive, especially the acquisition of various PCWDEs by plant-parasitic nematodes [3] and the acquisition of a mannanase in the coffee berry borer [23]. It is also clear that possessing enzymes able to degrade the hemicellulose network, such as GH11 xylanases, alongside enzymes targeting cellulose and pectins, provides an immediate advantage for a leaf beetle, giving it the ability to more efficiently process its food [24]. But the fact that no GH11 genes could be identified from other closely related leaf beetle species [22] reinforces the idea that the HGT event we described here must probably initially occurred randomly in *P. cochleariae* itself or at least in a common ancestor of the genus *Phaedon*. After the HGT, the original gh11 gene was maintained and fixed in the population due to natural selection, and provided a sufficient advantage to the beetle to then start to expand by gene duplication. There is no doubt that the exponential increase of animal genomes and transcriptomes being sequenced nowadays will reveal even more examples of functional micro-organism-to-animal HGTs. We believe that this will be particularly true in insects of the extraordinarily diverse order Coleoptera.

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch for technical support and Roy Kirsch for making available to us the *P. cochleariae* fosmid library. We thank Emily Wheeler, Boston, for editorial assistance. This work was supported by the Max Planck Gesellschaft.

### References


