Received: 24 May 2023 | Revised: 11 July 2023 | Accepted: 13 July 2023

DOI: 10.1002/arch.22039

RESEARCH ARTICLE

Archives of INSECT BIOCHEMISTRY AND PHYSIOLOGY WILEY

First evidence of a horizontally-acquired GH-7 cellobiohydrolase from a longhorned beetle genome

Na Ra Shin | Yannick Pauchet

Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, Jena, Germany

Correspondence
Yannick Pauchet, Insect Symbiosis, Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany.
Email: ypauchet@ice.mpg.de

Present address
Na Ra Shin, Department of Biological Sciences and Center for Biodiversity Research, University of Memphis, Memphis, Tennessee, USA.

Funding information
Max Planck Society; Deutsche Forschungsgemeinschaft, Grant/Award Number: PA2808/4-1

Abstract

Xylophagous larvae of longhorned beetles (Coleoptera; Cerambycidae) efficiently break down polysaccharides of the plant cell wall, which make the bulk of their food, using a range of carbohydrate-active enzymes (CAZymes). In this study, we investigated the function and evolutionary history of the first identified example of insect-encoded members of glycoside hydrolase family 7 (GH7) derived from the Lamiinae *Exocentrus adspersus*. The genome of this beetle contained two genes encoding GH7 proteins located in tandem and flanked by transposable elements. Phylogenetic analysis revealed that the GH7 sequences of *E. adspersus* were closely related to those of Ascomycete fungi, suggesting that they were acquired through horizontal gene transfer (HGT) from fungi. However, they were more distantly related to those encoded by genomes of Crustacea and of protist symbionts of termites and cockroaches, supporting that the same enzyme family was recruited several times independently in Metazoa during the course of their evolution. The recombinant *E. adspersus* GH7 was found to primarily break down cellulose polysaccharides into cellobiose, indicating that it is a cellobiohydrolase, and could also use smaller cellulose oligomers as substrates. Additionally, the cellobiohydrolase

activity was boosted by the presence of calcium chloride. Our findings suggest that the combination of GH7 cellobiohydrolases with other previously characterized endo-β−1,4-glucanases and β-glucosidases allows long-horned beetles like *E. adspersus* to efficiently break down cellulose into monomeric glucose.

KEYWORDS

Cerambycidae, digestive enzyme, gene duplication, glycoside hydrolase, horizontal gene transfer, longhorned beetles

Highlights

- Cellobiohydrolases of glycoside hydrolase family 7 (GH7) in longhorned beetles likely came from Ascomycete fungus via HGT.
- GH7 cellobiohydrolases were independently acquired multiple times in Metazoa through horizontal gene transfer.
- GH7 cellobiohydrolases, along with other enzymes, help longhorned beetle larvae break down cellulose into glucose units in their digestive tract.

1 | INTRODUCTION

Beetles of the family Cerambycidae, commonly known as longhorned beetles, is one of the largest beetle families, with over 36,000 extant species. This family of beetles is widely distributed, with species found in most parts of the world. They are particularly abundant in tropical and subtropical regions, but can also be found in temperate and boreal regions (Monné et al., 2017). Longhorned beetles are typically wood-boring insects, with larvae that tunnel into various types of wood, including trees, shrubs, and dead wood. They are considered both beneficial and harmful, as some species are important for pollination, and others can damage wood products, including timber and furniture. Some species, such as the Asian longhorned beetle (*Anoplophora glabripennis*, Lamiinae), have become serious invasive pest species because they develop into healthy trees, including maple, birch, elm, and willow, causing important economic damage in urban and forest areas. Other species also play a significant role in forest ecosystems as decomposers, helping to break down dead wood, thus participating nutrient recycling (Haack, 2017; Hanks, 1999; Linsley, 1959; Svacha & Lawrence, 2014).

Due to their xylophagous nature, larvae of Cerambycidae have to ingest large amounts of woody material to extract enough nutrients to develop and survive. The bulk of wood is made of the plant cell wall, which is a rigid layer that provides structural support and protection for the plant cell. It is composed of cellulose, hemicellulose, pectins, and lignin, which are complex polysaccharides and aromatic compounds that form a strong, fibrous network. This network is resistant to mechanical and enzymatic degradation and provides protection against pathogens and other environmental stresses (Doblin et al., 2010). There is a growing body of evidence that plant cell wall-degrading enzymes (PCWDEs) play an important role in the ability of cerambycid larvae to digest and

utilize the complex polysaccharides found in the plant cell wall (McKenna et al., 2016; Pauchet et al., 2014; Scully et al., 2013). Our recent work contributed to identify several types of PCWDEs produced by these insects targeting cellulose, hemicellulose, and pectins that we believe to be essential for their ability to feed on and digest wood (Shin et al., 2021). All of these carbohydrate-active enzymes (CAZymes) are members of several families of glycoside hydrolases (GHs). Among these CAZymes, we recently showed the pivotal role played by endo-acting enzymes of subfamily 2 of GH5 that evolved, after several gene duplication events, to break down not only cellulose but also xyloglucan, xylan, and mannans, the main constituents of the plant cell wall (Pauchet et al., 2020; Shin et al., 2022).

Our recent analyses of transcriptomes and genomes of more than two dozen species of Cerambycidae identified the first example in an insect of genes encoding proteins of family 7 of GHs (GH7) in the Lamiinae *Exocentrus adspersus* (Shin et al., 2023; Shin et al., 2021). GH7 enzymes are involved in the breakdown of complex carbohydrates, specifically cellulose and hemicellulose (Gilkes et al., 1991; Henrissat et al., 1989). These enzymes are found in a variety of organisms, including mostly microorganisms (Momeni et al., 2013; Ståhlberg et al., 1996), but also in animals, specifically Crustacea (Kern et al., 2013), in symbiotic protists of termites and cockroaches (Sethi et al., 2013; Watanabe et al., 2002), and in symbiotic *Termitomyces* fungal gardens maintained by higher termites from the genus Macrotermes (Poulsen et al., 2014). Several GH7 cellulases have been characterized and can be classified as being either endo-β−1,4- or exo-β−1,4-glucanases. The latter type of enzymes were shown to release cellobiose from cellulose polymers, indicating that they were cellobiohydrolases. Both types of GH7 cellulases can be discriminated based on their amino acid sequence due to the absence/presence of so-called tunnel-forming loops (Sethi et al., 2013). Finally, GH7 cellulases are classified as retaining enzymes, meaning that the stereochemistry of the bond is retained during the cleavage reaction (Knowles et al., 1988).

Here, we examined the function and the evolutionary history of GH7 cellulases encoded by the genome of the Lamiinae *E. adspersus*. Using phylogenetic analyses, we show that these insect-derived GH7 proteins are closely related to their counterparts found in species of Pezizomycotina fungi, but are far related to other animal-encoded counterparts. We suggest that the common ancestor of several species of Lamiinae acquired GH7-encoding genes from a Pezizomycotina fungus donor through a horizontal gene transfer (HGT) event. Functional characterization of a recombinant form of the *E. adspersus* GH7 indicates that this enzyme is a cellobiohydrolase, mainly releasing cellobiose from cellulose poly- and oligosaccharides, and that its activity is sensitive to calcium chloride. We propose that this enzyme acts in concert with previously characterized cerambycid-derived endoglucanases and β-glucosidases to break down cellulose into glucose units, thus providing sugar nutrients to the insect.

2 | MATERIALS AND METHODS

2.1 | Sequence analysis

The GH7-encoding gene sequence was obtained from two sources: a larval midgut transcriptome of *E. adspersus* (Shin et al., 2021), and the draft genome of the same species (Shin et al., 2023). The sequences were manually curated and submitted to NCBI (accessions: OQ721898 and OQ721899). The curated nucleotide sequences were translated into amino acid sequences using the Translate tool available on the Expasy bioinformatics resource portal of the Swiss Institute of Bioinformatics (Duvaud et al., 2021). SignalP v6.0 (Teufel et al., 2022) was used to identify a predicted signal peptide for secretion. To assess the presence of conserved catalytic residues, the amino acid sequence of *E. adspersus* GH7-1 was first analyzed using InterProScan (Paysan-Lafosse et al., 2023). Additionally, a BLASTp search was conducted against the Protein Data Bank (PDB) using the *E. adspersus* GH7-1 sequence as a query. Amino acid sequences of GH7 derived from fungi and Crustacea that showed high sequence similarities with the *E. adspersus* sequence, and for which information about the identity of catalytic residues was available, were recovered and aligned using MUSCLE v3.7 implemented in MEGA7 (Kumar et al., 2016).

2.2 | Phylogenetic analysis

Amino acid sequences of GH7 proteins derived from microorganisms and animals were recovered from public databases. Amino acid alignments were carried out using MAFFT v7.471 (Kuraku et al., 2013), and were inspected and corrected manually when needed. Phylogenetic relationships were estimated using maximum likelihood (ML) analysis and were conducted using the IQ-TREE web server (Trifinopoulos et al., 2016). The best model of protein evolution was determined within the IQ-TREE web server using ModelFinder (Kalyaanamoorthy et al., 2017). The reliability of each analysis was assessed by UFboot (Minh et al., 2013), which conducted an ultrafast bootstrap approximation with 1,000 replicates. Sequence alignment and tree file (Newick format) can be found in Supporting Information Data S1.

2.3 | Heterologous expression in insect *Sf9* cells

The procedure used in this study was similar to that described previously (Pauchet et al., 2020). In brief, we amplified the open reading frame (ORF) of *E. adspersus* GH7-1, excluding the stop codon, using polymerase chain reaction (PCR) with RACE-ready cDNA generated in a previous study (Shin et al., 2021). A Kozak sequence was added to the 5′-end of the PCR product by integrating it into the forward PCR primer (Table S1). The resulting PCR products were cloned into pIB/V5-His TOPO/TA (Invitrogen, Waltham, MA, USA) in a frame with a V5-$(His)_6$ epitope at the carboxyl-terminus, and colonies with constructs in the correct orientation were selected after colony PCR. Insect *Sf9* cells (Invitrogen) were cultured in SF-900 II serum-free medium (Gibco) and transfected with FUGENE HD (Promega) as the transfection reagent in 6-well plates. After 72 h, the culture medium was harvested, and recombinant GH7-1 proteins were recovered by immunoprecipitation using anti-V5 agarose beads (V5-Trap, ChromoTek). The agarose beads were then resuspended in 150 μL of double distilled water. Western blots were used to confirm the successful heterologous expression and pull-down of GH7-1, using an anti-V5 HRP antibody (Bethyl) and detecting the signal with the Radiance Plus Femtogram HRP substrate (Azure Biosystems) on an Azure 600 Imager (Azure Biosystems).

2.4 | Enzyme assays and thin-layer chromatography

To test the enzymatic activity, GH7-1 proteins bound to anti-V5 agarose beads were incubated with various cellulose-related substrates. A 20 μL enzyme assay was prepared by mixing the agarose beads (resuspended in 14 μL water) with a 1% substrate solution (4 μL) in a 20 mM citrate/phosphate buffer at pH 7.0. The substrates tested included carboxymethyl cellulose (CMC), regenerated amorphous cellulose (RAC), and microcrystalline cellulose (MCC). RAC was prepared following the method described in (Busch et al., 2018). Oligosaccharides purchased from Megazyme were also used as substrates. For these assays, 0.5 μL of a given oligosaccharide (10 μg/μL) was mixed with GH7-1 proteins bound to anti-V5 agarose beads (14 μL) in the same buffer. GH7-1 has tested with the following oligosaccharides: cellobiose to cellohexose. To test the effects of calcium chloride on the enzymatic activity of GH7-1, various concentrations ranging from 1 to 10 mM were added to the assays. After incubating the enzyme assays for either 24 or 72 h at 40°C, they were applied to thin-layer chromatography (TLC) plates (silica gel 60, 20 × 20 cm, Merck). The TLC plates were developed for a minimum of 180 min using a mobile phase composed of ethyl acetate/acetic acid/formic acid/water in a 9:3:1:4 ratio and then dried at room temperature. The hydrolysis products were then revealed by briefly heating the plates after soaking them in a solution of 0.2% (w/v) orcinol in methanol/sulfuric acid (9:1) until spots appeared. Substrates incubated in the above-described assay conditions, but without agarose beads coated with GH7-1, were included as controls.

Agarose beads coated with GH7-1, incubated in the assay conditions described above, but without any substrate, were also included as controls.

3 | RESULTS

3.1 | The genome of *E. adspersus* harbors two GH7 genes

Surveying the larval gut transcriptome of the Lamiinae *E. adspersus* (Shin et al., 2021), we initially identified a transcript with a 1347 bp-long ORF encoding a protein of 448 amino acids sharing up to 69% sequence identity with fungal glycoside hydrolase family 7 (GH7) proteins. We found the corresponding gene located on a 130-kb contig (Genbank accession: JANEYG010000319.1) of a draft genome of the same species (Shin et al., 2023). The only other gene present on this contig encoded another GH7 protein, and both genes were surrounded by insect-like transposable elements. We identified both class I (retrotransposons) and class II (DNA transposons) as well as simple repeats in the genomic regions surrounded both GH7-encoded genes (Table S2). We named the products of these two genes GH7-1 and −2. The ORFs of these two genes differ by only three nucleotide substitutions, one of them being nonsynonymous, and consequently the corresponding proteins differ by only a single amino acid. Both genes contain a 53-bp intron with 100% identical sequences.

Surprisingly, by performing BLAST searches, we discovered a previously unreported GH7 sequence (Genbank accession: AMD09875.1) annotated as a β−1,4-exoglucanase, from *Mesosa myops*, another species of Lamiinae like *E. adspersus*. This suggests that GH7 proteins may be more widespread in Lamiinae than we initially presumed. All three Lamiinae-derived GH7 proteins contain a 20-amino-acid signal peptide for secretion, and catalytic and substrate-binding amino acids described for fungal GH7 cellobiohydrolases counterparts are conserved (Figure 1). Altogether, we hypothesized that GH7 proteins of *E. adspersus* and *M. myops* may be cellobiohydrolases.

3.2 | Several independent horizontal acquisitions of GH7-encoding genes occurred in animals

Proteins of the GH7 family are found in a broad range of fungi, including Ascomycetes and Basidiomycetes. However, they have also been discovered in several species of Crustacea and in symbiotic protists of termites and cockroaches. To gain insight into the evolutionary origins of GH7 family members found in Cerambycidae and other lineages such as Ascomycete and Basidiomycete fungi, Crustacea, and symbiotic protists of termites and cockroaches, we performed a maximum likelihood phylogenetic analysis (Figure 2). The GH7 sequences obtained from cerambycids are grouped together in a well-supported clade along with Pezizomycotina sequences, which confirms our previous findings (Shin et al., 2021). The GH7 sequences from Crustacea form a distinct monophyletic clade, located at the base of the unrooted tree and distantly related to the cerambycid-derived ones. The GH7 sequences obtained from symbiotic protists form a separate monophyletic clade, which is distantly related to GH7s of Crustacea and Cerambycidae sequences (Figure 2). Our analyses suggest that GH7 proteins have been acquired separately at least twice in Metazoa, and also in symbiotic protists present in termites and cockroaches.

3.3 | The *E. Adspersus* GH7 is a cellobiohydrolase sensitive to the presence of calcium chloride

As it was nearly impossible to design primers that could distinguish between both transcripts of *E. adspersus gh7* genes, we could only amplify *gh7-1* for heterologous expression in insect cells via PCR. Following expression in *Sf9*

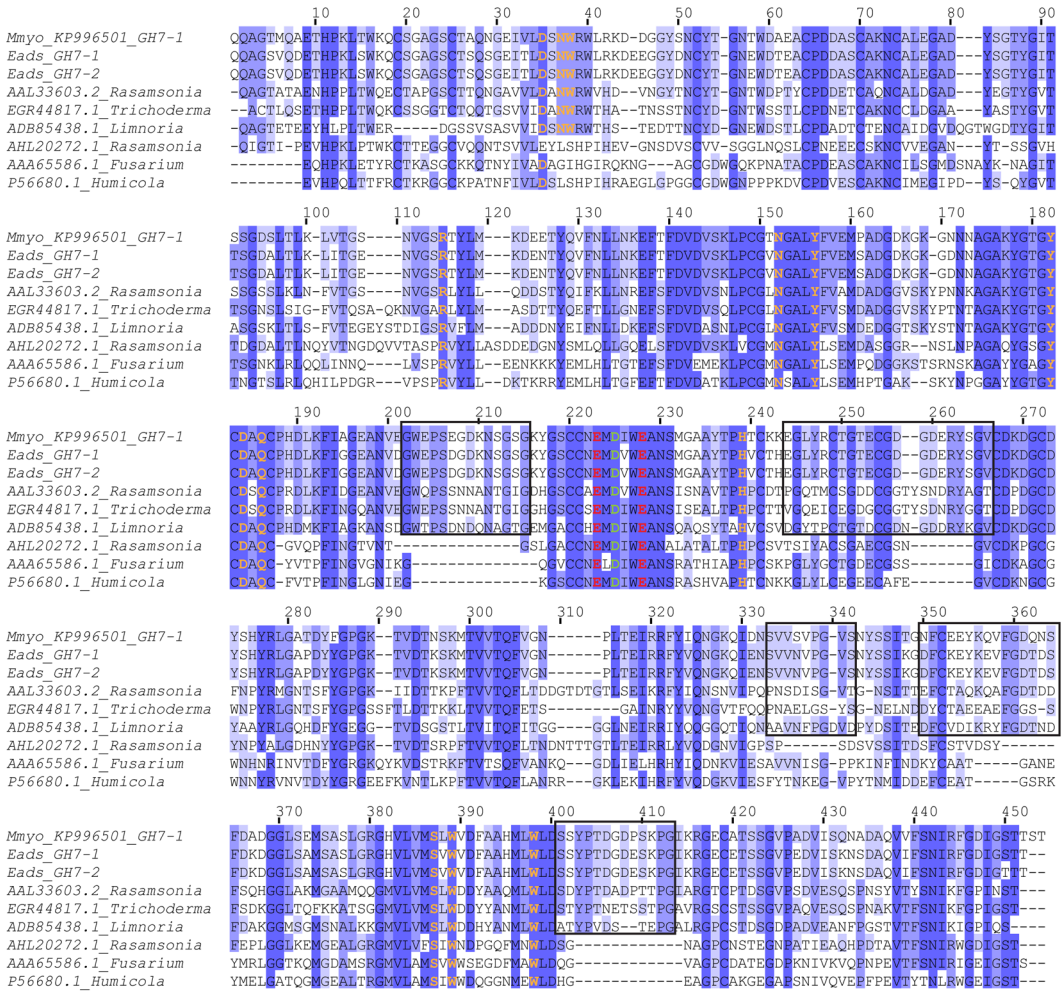

FIGURE 1 Lamiinae-derived GH7 proteins share high degree of similarities with fungal and Crustacea cellobiohydrolases. Multiple sequence alignment of GH7 proteins derived from the two lamiine *Exocentrus adspersus* (Genbank accessions: OQ721898 and OQ721899) and *Mesosa myops* (KP996501.1), the fungi *Trichoderma reesei* (EGR44817.1), *Fusarium oxysporum* (AAA65586.1), *Humicola insolens* (P56680.1) and *Rasamsonia emersonii* (AAL33603.2 and AHL20272.1), and the isopod *Limnoria quadripunctata* (ADB85438.1). The consensus motif -Glu-X-Asp-X-X-Glu- is conserved in the Lamiinae-derived sequence. The initial Glu residue functions as the catalytic nucleophile, while the other Glu residue acts as a general acid/base (both indicated in red). The function of the Asp residue of this motif (in green) is still unclear, but this residue was shown to be implicated in the catalytic reaction. The residues colored in orange were shown to interact with the substrate within the catalytic site of the enzyme. Boxes enclose predicted tunnel-forming loops in GH7 cellobiohydrolases.

cells, we were able to extract the recombinant GH7-1 protein that we pulled down using anti-V5 beads (Figure S1). Our initial enzyme assays, conducted using cellulose poly- and oligosaccharides as substrates under standard conditions (Shin et al., 2022), yielded disappointing results as the breakdown products were barely visible on TLC, even after long incubation times up to 72 h (data not shown). In addition, despite extensive washing of the beads, a monomeric sugar, which migrated at a similar distance as glucose, in our pulled-down enzyme samples was consistently observed on TLC (Figures 3 and 4). Based on our observations, we strongly suspect that this monomeric sugar is the primary carbon source present in the culture medium utilized for our *Sf*9 cell culture.

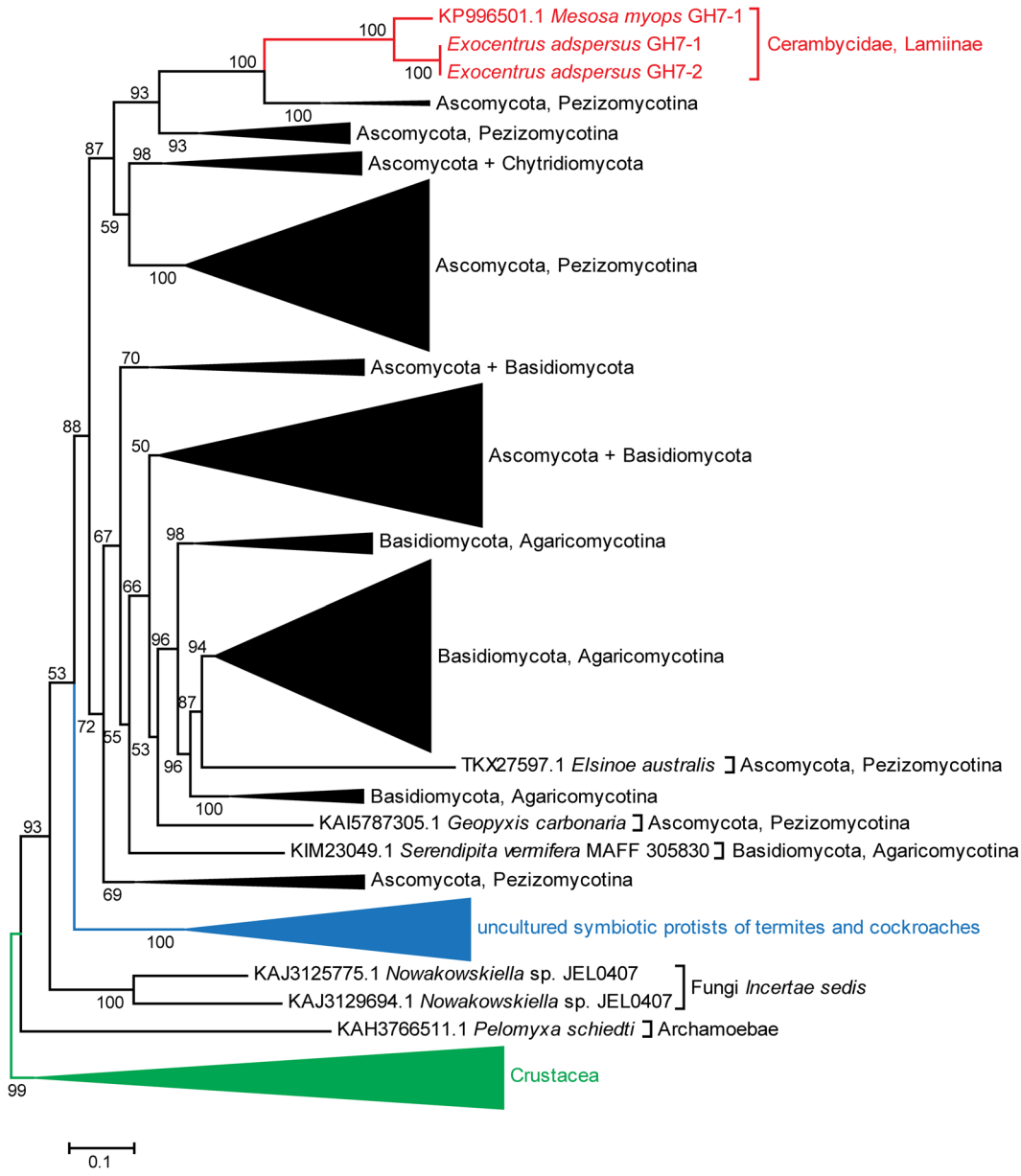

FIGURE 2 Phylogenetic relationships between GH7 family members identified from species of Cerambycidae and their animal and microbial counterparts. A maximum likelihood analysis was performed with 1000 UltraFast bootstrap replicates using IQ-TREE. The best model of protein evolution was determined in IQ-TREE and was the 'Whelan and Goldman' (WAG) model, with amino acid frequencies randomly generated from a uniform distribution (+F), and incorporating a FreeRate model (+R; number of categories = 9). UltraFast Bootstrap values are indicated next to the branches. The tree presented is rooted at midpoint. The original tree and the corresponding amino acid sequence alignment are provided as Supporting Information.

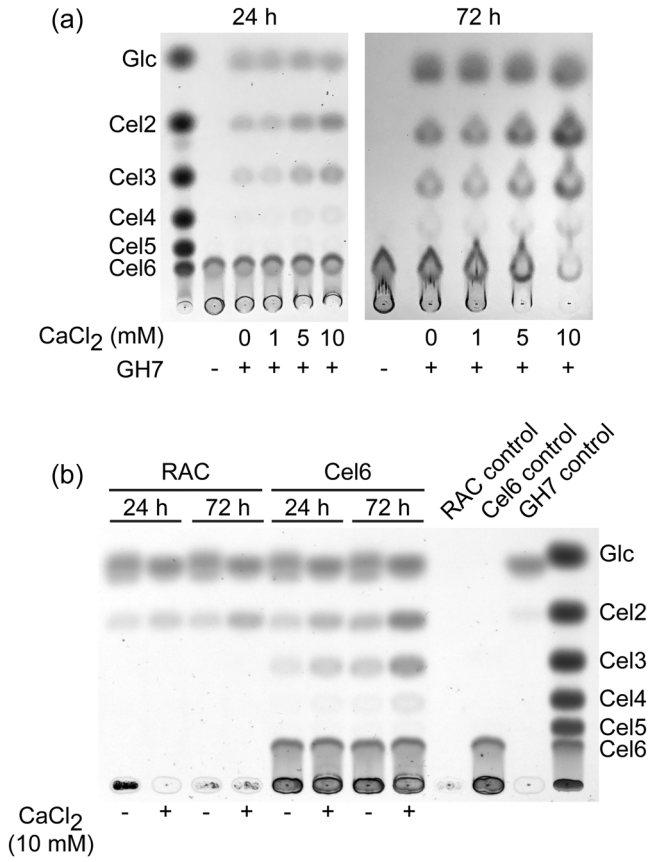

FIGURE 3 Calcium chloride boosts the enzymatic activity of *Exocentrus adspersus* GH7. (a) The activity of recombinant *E. adspersus* GH7 was tested with cellohexaose (Cel6) as a substrate, in the presence of increasing concentrations (1, 5, and 10 mM) of calcium chloride ($CaCl_2$) and at two different time points (24 and 72 h), in McIlvaine buffer pH 7.0 at 40°C. Breakdown products were analyzed by Thin-layer chromatography (TLC). The standards were glucose (Glc) to cellohexaose (Cel6). (b) Comparison of the breakdown products generated by the recombinant *E. adspersus* GH7 using either regenerated amorphous cellulose (RAC) or cellohexaose (Cel6) as substrates, in the absence or presence of 10 mM $CaCl_2$, and at two different time points (24 and 72 h). Substrate controls in the absence of enzyme ("RAC control" and "Cel6 control"), as well as enzyme control in the absence of substrate ("GH7 control") were also run on the TLC. The standards are the same as in (a).

It appears that this monomeric sugar may bind to the enzyme, likely in the active site, and is not easily displaced by simple washing procedures.

Considering the luminal pH of Cerambycidae larvae was found to be neutral to slightly alkaline in prior studies (Kukor & Martin, 1986; Shaw & Christeller, 2009), we opted to adjust the pH of our enzyme assays from 5.0 to 7.0. Under these conditions, we were able to detect breakdown products on TLC (as depicted in Figure 3). When incubated with regenerated amorphous cellulose (RAC), GH7-1 appeared to produce cellobiose as its main breakdown product, whereas cellohexaose as a substrate resulted in the detection of cellotetraose, cellotriose, and cellobiose on TLC. The release of cellobiose from a cellulose polysaccharide suggests that GH7-1 is most likely a cellobiohydrolase.

According to prior research, calcium ions have been shown to boost the enzymatic activity of GH7 cellobiohydrolases, especially those of termite symbionts (Sethi et al., 2013). Therefore, we investigated the impact of increasing concentrations of calcium chloride in our enzyme assays (Figure 3a). As the concentration of calcium

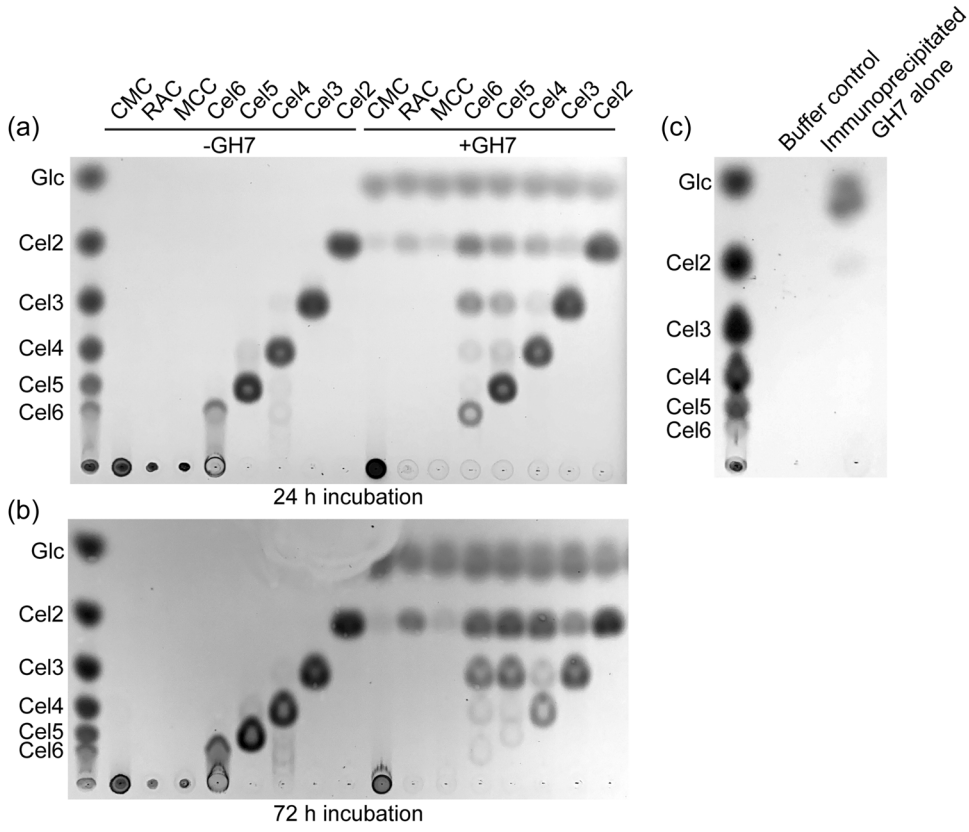

FIGURE 4 *Exocentrus adspersus* GH7 is a cellobiohydrolases. (a) Thin-layer chromatography (TLC) analyses of hydrolysis end-products of various cellulose poly- and oligosaccharides by *E. adspersus* GH7. Recombinant GH7, expressed in *Sf9* cells, was immunoprecipitated using anti-V5 agarose beads. Beads harboring the recombinant GH7 protein were incubated with carboxymethylcellulose (CMC), regenerated amorphous cellulose (RAC), microcrystalline cellulose (MCC), as well as, cellulose oligomers ranging from cellobiose (Cel2) to cellohexaose (Cel6) in McIlvaine buffer pH 7.0 containing 10 mM calcium chloride ($CaCl_{2)}$ at 40°C for 24 h before being spotted on TLC plates. The standards were glucose (Glc) to cellohexaose (Cel6). (b) Same experiments as described above, but incubated for 72 h. (c) TLC showing the presence of a monosaccharide migrating similarly to glucose in the mixture of agarose beads coated with the recombinant *E. adspersus* GH7 even after extensive washing.

chloride present in the assays increased, the breakdown products released by GH7-1 using cellohexaose as a substrate appeared more intense on TLC plates. After incubating for 72 h using 10 mM calcium chloride in our assays, a significant decrease in the amount of cellohexaose was observed alongside a noticeable increase in breakdown products (Figure 3a). Similar outcomes were obtained when RAC was used as a substrate, with a visible increase of cellobiose produced by GH7-1 under the same assay conditions as those described before (Figure 3b).

We further tested other cellulose poly- and oligosaccharides as potential substrates for GH7-1 (Figure 4). *Exocentrus adspersus* GH7-1 generated only minimal amounts of cellobiose when carboxymethylcellulose and microcrystalline cellulose were used as substrates, as opposed to regenerated amorphous cellulose, following incubation for 24 (Figure 4a) and 72 h (Figure 4b). Aside from its ability to use cellohexaose as a substrate, GH7-1 was also able to generate breakdown products from cellopentaose, cellotetraose, and, to a certain extent, cellotriose (Figure 4). Overall, GH7-1 from *E. adspersus* is a new insect-derived cellobiohydrolase that primarily generates cellobiose from cellulose polymers and can also utilize smaller cellulose oligomers as substrates.

4 | DISCUSSION

The efficient breakdown of plant cell wall polysaccharides by xylophagous larvae of longhorned beetles is a complex process that involves a range of carbohydrate-active enzymes (CAZymes). In this study, we focused on the function and evolutionary history of the first identified insect-encoded members of glycoside hydrolase family 7 (GH7) derived from the Lamiinae *Exocentrus adspersus*. Our results support the hypothesis that the *gh7* genes of *E. adspersus* were acquired through horizontal gene transfer (HGT) from fungi, but were also independently recruited in other Animal lineages during evolution. We found that the recombinant *E. adspersus* GH7 primarily acts as a cellobiohydrolase, releasing cellobiose from cellulose polysaccharides, and that its activity is enhanced by the presence of calcium chloride. Our findings shed new light on the intricate mechanisms that allow longhorned beetles to break down cellulose. Our study highlights the apparent significance of horizontal gene transfer (HGT) in the evolution of carbohydrate-active enzymes (CAZymes) in insects. Specifically, we propose that the GH7 cellobiohydrolases, which were horizontally acquired and act in conjunction with GH5 and GH45 endoglucanases (Pauchet et al., 2014; Shin et al., 2022), also previously characterized as horizontally acquired, as well as widely distributed GH1 β-glucosidases (McKenna et al., 2016, 2019), enable longhorned beetles to break down cellulose into glucose monomers.

By searching public sequence repositories, we have discovered another GH7 sequence in the Lamiinae *Mesosa myops*. Considering that this subfamily has the highest number of species within the Cerambycidae family (approximately 21,000 species) (Monné et al., 2017), we anticipate that with the increasing availability of transcriptomes and genomes for this beetle lineage, more genes encoding GH7 family members will be discovered in the future. Interestingly, by screening our *E. adspersus* draft genome, we found duplicated copies of the *gh7* gene located in tandem on the same contig and surrounded by insect-derived transposable elements. Taking into account that both ORFs were almost identical resulting in a single amino acid difference between the corresponding proteins located close to the C-terminal end, we propose that the duplication event occurred to increase the gene dosage, rather than for sub-functionalization or neofunctionalization purposes.

While some glycoside hydrolases have been reported to utilize cofactors such as metal ions or nicotinamide adenine dinucleotide (NAD) (Rajan et al., 2004), this is not a widespread characteristic of these enzymes. Therefore, the observation that the activity of GH7-1 is increased by the presence of calcium chloride is difficult to explain. Such a phenomenon has been previously described for GH7 cellobiohydrolases expressed by symbiont protists of termites (Sethi et al., 2013). It is yet to be determined if a calcium ion binding site is present in the active site of these enzymes, as no such site has been identified in GH7 microbial counterparts with known three-dimensional structures (Kern et al., 2013; Momeni et al., 2013; Ståhlberg et al., 1996). Therefore, it is possible that the calcium chloride enhancement effect could be attributed to the chloride ions rather than the calcium ions (Walker & Whelan, 1960), or to an interaction between the calcium chloride and the substrate rather than the enzyme (Tokuyasu et al., 2008).

Longhorned beetle larvae have two reasons to digest cellulose that are not mutually exclusive. First, breaking down cellulose down to simple glucose can provide a carbon source for these insects, and can be an alternative to starch and other sugars. Holl et al. (2002) conducted a study comparing the soluble sugar composition between *Pinus sylvestris* sapwood and the feces of *Hylotrupes bajulus* larvae (Cerambycidae; Cerambycinae), which utilize the former material as a food source. One of the main conclusions drawn from the analysis was that the levels of α-glucans (starch) in the feces were not significantly different from those found in Scots pine sapwood, suggesting that starch may not play a significant role in the larvae's nutrition (Holl et al., 2002). It can be speculated in this context that the breakdown of cellulose may compensate for the non-utilization of starch in cerambycid larvae. Secondly, breaking down cellulose and other plant cell wall components can make it easier to access the nutrient-rich content of plant cells. Kirsch et al. (2022) used CRISPR/Cas9 knockouts of pectinases to demonstrate that breaking down pectin, a major polysaccharide of the plant cell wall, was crucial for the survival of a leaf beetle (Kirsch et al., 2022). Their results showed that providing pectin breakdown products to the mutant leaf beetles did

not rescue them while providing pectinase activity did. Based on these findings, they concluded that leaf beetles did not digest pectin to acquire a source of energy, but to access the nutritious content of plant cells.

Due to the limited occurrence of GH7 cellobiohydrolases in Cerambycidae and other Phytophaga clade lineages, we questioned if their absence could be compensated by the presence of cellobiohydrolases from other glycoside hydrolase families. We have reasons to believe this is the case. Firstly, gut lumen samples from the Green Dock beetle *Gastrophysa viridula* have been shown to break down cellulose poly- and oligosaccharides into simple glucose units, suggesting the concerted action of endoglucanases, cellobiohydrolases, and β-glucosidases (Busch et al., 2018). Secondly, bacterial counterparts of one of the most widely distributed CAZyme families in Phytophaga beetles, GH48, have been characterized as cellobiohydrolases (McKenna et al., 2019; Pauchet et al., 2010; Shin et al., 2021; Sukharnikov et al., 2012). Intriguingly, the genome of *E. adspersus* lacks genes encoding putative GH48 family members, while other Cerambycidae species have them (McKenna et al., 2019; Shin et al., 2023; Shin et al., 2021). However, the sole functionally characterized GH48 family member of the leaf beetle *Gastrophysa atrocyanea* seems to be active on chitin (Fujita et al., 2006), which is puzzling since genes encoding the classical chitinase family GH18 are widely distributed in insects in general (Chen et al., 2020). In this context, the enzymatic activity of beetle-derived GH48 family members warrants re-examination in the future, as they may indeed encompass cellobiohydrolases.

AUTHOR CONTRIBUTIONS

Na Ra Shin: Conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing—review & editing. Yannick Pauchet: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; validation; visualization; writing—original draft; writing—review & editing. All authors edited the manuscript.

ACKNOWLEDGMENTS

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch for their technical support. This work was supported by the Max Planck Society. Yannick Pauchet acknowledges support from the Deutsche Forschungsgemeinschaft (DFG; PA2808/4-1). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict to disclose.

DATA AVAILABILITY STATEMENT

Supplementary tables (Tables S1 and S2) and figures (Figure S1), as well as the maximum likelihood phylogenetic analysis (Newick format) and the corresponding sequence alignment (Fasta format) have been submitted with the present paper and are available online. The sequences of both GH7 have been submitted to NCBI Genbank. The data that supports the findings of this study are available in the supplementary material of this article

ORCID

Yannick Pauchet http://orcid.org/0000-0002-2918-8946

REFERENCES

Busch, A., Kunert, G., Wielsch, N. & Pauchet, Y. (2018) Cellulose degradation in *Gastrophysa viridula* (Coleoptera: Chrysomelidae): Functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity. *Insect Molecular Biology*, 27(5), 633–650. Available from: https://doi.org/10.1111/imb.12500

Chen, W., Jiang, X. & Yang, Q. (2020) Glycoside hydrolase family 18 chitinases: the known and the unknown. *Biotechnology Advances*, 43, 107553. Available from: https://doi.org/10.1016/j.biotechadv.2020.107553

Doblin, M.S., Pettolino, F. & Bacic, A. (2010) Plant cell walls: the skeleton of the plant world. *Functional Plant Biology*, 37(5), 357–381. Available from: https://doi.org/10.1071/FP09279

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V. & Durinx, C. (2021) Expasy, the Swiss bioinformatics resource portal, as designed by its users. *Nucleic Acids Research*, 49(W1), W216–W227. Available from: https://doi.org/10.1093/nar/gkab225

Fujita, K., Shimomura, K., Yamamoto, K., Yamashita, T. & Suzuki, K. (2006) A chitinase structurally related to the glycoside hydrolase family 48 is indispensable for the hormonally induced diapause termination in a beetle. *Biochemical and Biophysical Research Communications*, 345(1), 502–507. Available from: https://doi.org/10.1016/j.bbrc.2006.04.126

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, Jr. R.C. & Warren, R.A. (1991) Domains in microbial beta-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiological Reviews*, 55(2), 303–315. Available from: https://doi.org/10.1128/mr.55.2.303-315.1991

Haack, R.A. (2017) Feeding biology of Cerambycids. In: Wang, Q. (Ed.) *Cerambycidae of the world: biology and pest management*. Boca Raton, FL: CRC Press. pp. 105–132.

Hanks, L.M. (1999) Influence of the larval host plant on reproductive strategies of cerambycid beetles. *Annual Review of Entomology*, 44, 483–505. Available from: https://doi.org/10.1146/annurev.ento.44.1.483

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L. & Mornon, J.P. (1989) Cellulase families revealed by hydrophobic cluster analysi. *Gene*, 81(1), 83–95. Available from: https://doi.org/10.1016/0378-1119(89)90339-9

Holl, W., Frommberger, M. & Strassl, C. (2002) Soluble carbohydrates in the nutrition of house longhorn beetle larvae, *Hylotrupes bajulus* (L.) (Col., Cerambycidae): from living sapwood to faeces. *Journal of Applied Entomology*, 126(9), 463–469. Available from: https://doi.org/10.1046/j.1439-0418.2002.00696.x

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. Available from: https://doi.org/10.1038/nmeth.4285

Kern, M., McGeehan, J.E., Streeter, S.D., Martin, R.N.A., Besser, K., Elias, L. et al. (2013) Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences*, 110(25), 10189–10194. Available from: https://doi.org/10.1073/pnas.1301502110

Kirsch, R., Okamura, Y., Haeger, W., Vogel, H., Kunert, G. & Pauchet, Y. (2022) Metabolic novelty originating from horizontal gene transfer is essential for leaf beetle survival. *Proceedings of the National Academy of Sciences*, 119(40), e2205857119. Available from: https://doi.org/10.1073/pnas.2205857119

Knowles, J.K.C., Lentovaara, P., Murray, M. & Sinnott, M.L. (1988) Stereochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. *Journal of the Chemical Society, Chemical Communications*, 1401–1402. Available from: https://doi.org/10.1039/C39880001401

Kukor, J.J. & Martin, M.M. (1986) The transformation of *Saperda calcarata* (Coleoptera, Cerambycidae) into a cellulose digester through the inclusion of fungal enzymes in its diet. *Oecologia*, 71(1), 138–141. Available from: https://doi.org/10.1007/Bf00377333

Kumar, S., Stecher, G. & Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. Available from: https://doi.org/10.1093/molbev/msw054

Kuraku, S., Zmasek, C.M., Nishimura, O. & Katoh, K. (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, 41, W22–W28. Available from: https://doi.org/10.1093/nar/gkt389

Linsley, E.G. (1959) Ecology of cerambycidae. *Annual Review of Entomology*, 4, 99–138. Available from: https://doi.org/10.1146/annurev.en.04.010159.000531

McKenna, D.D., Scully, E.D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S.M. et al. (2016) Genome of the asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biology*, 17(1), 227. Available from: https://doi.org/10.1186/s13059-016-1088-8

McKenna, D.D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D.J. et al. (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences*, 116(49), 24729–24737. Available from: https://doi.org/10.1073/pnas.1909655116

Minh, B.Q., Nguyen, M.A.T. & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. Available from: https://doi.org/10.1093/molbev/mst024

Momeni, M.H., Payne, C.M., Hansson, H., Mikkelsen, N.E., Svedberg, J., Engström, Å. et al. (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. *Journal of Biological Chemistry*, 288(8), 5861–5872. Available from: https://doi.org/10.1074/jbc.M112.440891

Monné, M.L., Monné, M.A. & Wang, Q. (2017) General morphology, classification, and biology of Cerambycidae. In: Wang, Q. (Ed.) *Cerambycidae of the world: biology and pest management*. Boca Raton, FL: CRC Press. pp. 1–70.

Pauchet, Y., Kirsch, R., Giraud, S., Vogel, H. & Heckel, D.G. (2014) Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle *Apriona japonica*. *Insect Biochemistry and Molecular Biology*, 49, 1–13. Available from: https://doi.org/10.1016/j.ibmb.2014.03.004

Pauchet, Y., Ruprecht, C. & Pfrengle, F. (2020) Analyzing the substrate specificity of a class of long-horned-beetle-derived xylanases by using synthetic arabinoxylan oligo- and polysaccharides. *ChemBioChem*, 21(10), 1517–1525. Available from: https://doi.org/10.1002/cbic.201900687

Pauchet, Y., Wilkinson, P., Chauhan, R. & Ffrench-Constant, R.H. (2010) Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PLoS One*, 5(12), e15635. Available from: https://doi.org/10.1371/journal.pone.0015635

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A. et al. (2023) InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418–D427. Available from: https://doi.org/10.1093/nar/gkac993

Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S. et al. (2014) Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences*, 111(40), 14500–14505. Available from: https://doi.org/10.1073/pnas.1319718111

Rajan, S.S., Yang, X., Collart, F., Yip, V.L.Y., Withers, S.G., Varrot, A. et al. (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD+/Mn2+-dependent phospho-α-glucosidase from bacillus subtilis. *Structure*, 12(9), 1619–1629. Available from: https://doi.org/10.1016/j.str.2004.06.020

Scully, E.D., Hoover, K., Carlson, J.E., Tien, M. & Geib, S.M. (2013) Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle. *BMC Genomics*, 14(1), 850. Available from: https://doi.org/10.1186/1471-2164-14-850

Sethi, A., Kovaleva, E.S., Slack, J.M., Brown, S., Buchman, G.W. & Scharf, M.E. (2013) A GHF7 cellulase from the protist symbiont community of *Reticulitermes flavipes* enables more efficient lignocellulose processing by host enzymes. *Archives of Insect Biochemistry and Physiology*, 84(4), 175–193. Available from: https://doi.org/10.1002/arch.21135

Shaw, B.D. & Christeller, J.T. (2009) Characterization of the proteases in the midgut of the xylophagous larvae of *Oemona hirta* (Coleoptera: Cerambycidae). *Insect Science*, 16(5), 381–386. Available from: https://doi.org/10.1111/j.1744-7917.2009.01274.x

Shin, N.R., Doucet, D. & Pauchet, Y. (2022) Duplication of horizontally acquired GH5_2 enzymes played a central role in the evolution of longhorned beetles. *Molecular Biology and Evolution*, 39(6), msac128. Available from: https://doi.org/10.1093/molbev/msac128

Shin, N.R., Okamura, Y., Kirsch, R. & Pauchet, Y. (2023) Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles. *Insect Molecular Biology*. Available from: https://doi.org/10.1111/imb.12844

Shin, N.R., Shin, S., Okamura, Y., Kirsch, R., Lombard, V., Svacha, P. et al. (2021) Larvae of longhorned beetles (Coleoptera; Cerambycidae) have evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes. *Systematic Entomology*, 46(4), 784–797. Available from: https://doi.org/10.1111/syen.12488

Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T.T. et al. (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *Journal of Molecular Biology*, 264(2), 337–349. Available from: https://doi.org/10.1006/jmbi.1996.0644

Sukharnikov, L.O., Alahuhta, M., Brunecky, R., Upadhyay, A., Himmel, M.E., Lunin, V.V. et al. (2012) Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. *Journal of Biological Chemistry*, 287(49), 41068–41077. Available from: https://doi.org/10.1074/jbc.M112.405720

Svacha, P. & Lawrence, J.F. (2014) Cerambycidae Latreille, 1802. In: Beutel, R.G. & Leschen, R.A.B. (Eds.) *Handbook of zoology. Arthropoda: Insecta. Coleoptera, beetles: morphology and systematics (Phytophaga)*, 3. Berlin: Walter de Gruyter. pp. 77–178.

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D. et al. (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, 40(7), 1023–1025. Available from: https://doi.org/10.1038/s41587-021-01156-3

Tokuyasu, K., Tabuse, M., Miyamoto, M., Matsuki, J. & Yoza, K. (2008) Pretreatment of microcrystalline cellulose flakes with CaCl2 increases the surface area, and thus improves enzymatic saccharification. *Carbohydrate Research*, 343(7), 1232–1236. Available from: https://doi.org/10.1016/j.carres.2008.03.007

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A. & Minh, B.Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(W1), W232–W235. Available from: https://doi.org/10.1093/nar/gkw256

Walker, G.J. & Whelan, W.J. (1960) The mechanism of carbohydrase action. 7. *Biochemical Journal*, 76(2), 257–263. Available from: https://doi.org/10.1042/bj0760257

Watanabe, H., Nakashima, K., Saito, H. & Slaytor, M. (2002) New endo-β−1,4-glucanases from the parabasalian symbionts, *Pseudotrichonympha grassii* and *Holomastigotoides mirabile* of Coptotermes termites. *Cellular and Molecular Life Sciences*, 59(11), 1983–1992. Available from: https://doi.org/10.1007/PI00012520

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shin, N.R. & Pauchet, Y. (2023) First evidence of a horizontally-acquired GH-7 cellobiohydrolase from a longhorned beetle genome. *Archives of Insect Biochemistry and Physiology*, 114, e22039. https://doi.org/10.1002/arch.22039