


# Duplication of Horizontally Acquired GH5\_2 Enzymes Played a Central Role in the Evolution of Longhorned Beetles

Na Ra Shin,<sup>1,2</sup> Daniel Doucet,<sup>3</sup> and Yannick Pauchet <sup>\*,1,2</sup>

<sup>1</sup>Department of Entomology, Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany

<sup>2</sup>Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany

<sup>3</sup>Great Lakes Forestry Centre, Natural Resources Canada, Canadian Forest Service, Sault Ste. Marie, ON P6A 2E5, Canada

\*Corresponding author: E-mail: ypauchet@ice.mpg.de.

Associate Editor: Aya Takahashi

## Abstract

**The rise of functional diversity through gene duplication contributed to the adaptation of organisms to various environments. Here we investigate the evolution of putative cellulases of the subfamily 2 of glycoside hydrolase family 5 (GH5\_2) in the Cerambycidae (longhorned beetles), a megadiverse assemblage of mostly xylophagous beetles. Cerambycidae originally acquired GH5\_2 from a bacterial donor through horizontal gene transfer (HGT), and extant species harbor multiple copies that arose from gene duplication. We ask how these digestive enzymes contributed to the ability of these beetles to feed on wood. We analyzed 113 GH5\_2, including the functional characterization of 52 of them, derived from 25 species covering most subfamilies of Cerambycidae. Ancestral gene duplications led to five well-defined groups with distinct substrate specificity, allowing these beetles to break down, in addition to cellulose, polysaccharides that are abundant in plant cell walls (PCWs), namely, xyloglucan, xylan, and mannans. Resurrecting the ancestral enzyme originally acquired by HGT, we show it was a cellulase that was able to break down glucomannan and xylan. Finally, recent gene duplications further expanded the catalytic repertoire of cerambycid GH5\_2, giving rise to enzymes that favor transglycosylation over hydrolysis. We suggest that HGT and gene duplication, which shaped the evolution of GH5\_2, played a central role in the ability of cerambycid beetles to use a PCW-rich diet and may have contributed to their successful radiation.**

**Key words:** horizontal gene transfer, gene duplication, glycoside hydrolase, digestive enzyme, cerambycidae, longhorned beetles.

## Introduction

The longhorned beetle family Cerambycidae has an estimated 36,300 extant species (Monné *et al.* 2017); the immature larval stage of most species in this megadiverse clade of phytophagous insects is xylophagous (Linsley 1959; Svacha and Lawrence 2014). Larvae of longhorned beetles develop in a challenging environment, as they have to deal with large amounts of difficult-to-digest plant cell wall (PCW), which make up the bulk of their food (Hanks 1999; Haack 2017). To extract as many nutrients as possible from their diet, cerambycid larvae express a range of so-called PCW degrading enzymes (PCWDEs) in their digestive tract (Shin *et al.* 2021). Beetle larvae use these enzymes to break down PCW polysaccharides, thus facilitating access to the more nutrient-rich cytoplasm of plant cells. The genomes of Cerambycidae typically encode well-known predicted cellulolytic and pectolytic enzyme families, such as glycoside hydrolase (GH) families 9 (GH9), 45 (GH45), and 48 (GH48) cellulases as well as GH28 polygalacturonases (pectinases) (Shin *et al.* 2021). However, the apparent lack of PCWDE families known to break down abundant hemicellulose polysaccharides—

such as xyloglucan, xylan, and mannans—in most subfamilies of Cerambycidae is striking. Species of the subfamily Cerambycinae are an exception to this rule; indeed, beside cellulolytic and pectolytic enzyme families, their genomes encode GH5 subfamily 8 (GH5\_8) mannanases as well as GH43\_26 xylan-debranching enzymes (Shin *et al.* 2021). Whether species of Cerambycidae, apart from those of the subfamily Cerambycinae, can break down hemicellulose polysaccharides, particularly abundant in woody tissues, remains an open question.

A trademark of Cerambycidae genomes, in terms of the complement of PCWDE families they encode, is the presence of the cellulolytic GH5\_2 family (Scully *et al.* 2013; Pauchet *et al.* 2014; McKenna *et al.* 2016). Indeed, these enzymes are absent from the genomes of closely related species of leaf beetles (Chrysomelidae) and weevils (Curculionoidea) (McKenna *et al.* 2019; Shin *et al.* 2021). Independently, several studies have pointed out that the presence of GH5\_2 in Cerambycidae is due to a horizontal gene transfer (HGT) from a bacterial donor, likely a species of Bacteroidetes (Danchin *et al.* 2010; McKenna *et al.* 2019; Shin *et al.* 2021). Extant species of Cerambycidae are known to harbor several copies of GH5\_2, indicating that several gene duplications

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

occurred after the initial HGT event, giving rise to a moderately sized gene family (Shin et al. 2021). The first-ever cellulase identified in a beetle was a GH5\_2 endo- $\beta$ -1,4-glucanase from the yellow-spotted longicorn beetle *Psacotheta hilaris* (Cerambycidae) (Sugimura et al. 2003). What we know about the function of GH5\_2 enzymes in longhorned beetles is restricted to species of the subfamily Lamiinae. Most of the Lamiinae-derived GH5\_2 that have been characterized so far are endo- $\beta$ -1,4-glucanases mainly acting on amorphous cellulose (Sugimura et al. 2003; Wei et al. 2006; Chang et al. 2012). However, recent studies have indicated that, in several species of Lamiinae, paralogs evolved to break down other PCW polysaccharides, such as xylan and xyloglucan (Pauchet et al. 2014, 2020; McKenna et al. 2016). These data suggest that at least some species of Lamiinae may use GH5\_2 paralogs to digest hemicellulose polysaccharides.

Family 5 of glycoside hydrolase (GH5) is a large family of carbohydrate-active enzymes (CAZymes) (Drula et al. 2022). Historically, GH5 has been classified in a group of enzymes called 'cellulase family A (GH-A)' (Henrissat et al. 1989; Gilkes et al. 1991), but enzymes of this family have also been shown to act on polysaccharides other than cellulose (Aspeborg et al. 2012). GH5 enzymes usually hydrolyze  $\beta$ -linked glycosidic bonds from a range of oligo- and polysaccharides, using a retaining double-displacement mechanism (Barras et al. 1992) based on two glutamates as conserved catalytic residues (Henrissat et al. 1995; Jenkins et al. 1995). The GH5 family is so large that it has been further classified into 51 subfamilies according to similarities on amino acid level and phylogenetic relationships. Functional data available so far indicate that about one-third of these subfamilies are monospecific, that is, they contain a single enzyme activity (Aspeborg et al. 2012). In particular, subfamily 2 (GH5\_2), the largest subfamily within GH5, is composed of  $\beta$ -1,4-glucan cleaving enzymes, mostly cellulases (EC 3.2.1.4), derived predominantly from bacteria (Aspeborg et al. 2012). Recent studies have indicated that a few eukaryote lineages within Nematoda and Insecta also possess GH5\_2. These organisms, which include plant-parasitic nematodes (Danchin et al. 2010) and, as mentioned above, longhorned beetles (McKenna et al. 2016; Shin et al. 2021), use GH5\_2 enzymes either to help them invade plants and establish parasitism, or to digest their plant-based diet.

Gene duplication played an essential role in the evolution of novel functions (Innan and Kondrashov 2010). However, the fate of newly duplicated genes can be quite different. Although some duplicated genes rapidly accumulate deleterious mutations and are subsequently lost, others are preserved and keep the same function as the original gene, resulting in an increased gene dosage. The function of duplicated genes can also evolve by either subfunctionalization or neofunctionalization. Subfunctionalization results in a subdivision of gene functions among the duplicated genes that are inherited from the original gene. In the case of neofunctionalization, duplicated genes accumulate neutral mutations, resulting

in the random acquisition of a completely new function (Lynch and Force 2000; Walsh 2003; Innan and Kondrashov 2010; McGrath et al. 2014). In the evolution of enzymes, diverse environmental conditions have often resulted in gene duplications that have increased substrate specificity and catalytic promiscuity (David and Alm 2011).

The advent of new sequencing technologies, and the rapid accumulation of transcriptome and genome data, have greatly facilitated the study of gene duplication in general. However, large-scale functional analyses of the outcome of gene duplication are often missing. Consequently, the currently available functional data do not reflect the wide distribution of GH5\_2 within the family Cerambycidae (Shin et al. 2021) and fail to account for the diversity of this group of insects. To address this lack of knowledge, we analyzed the evolution of 113 GH5\_2 sequences recovered from our recent transcriptome analysis of 25 species, which represent six out of eight subfamilies of Cerambycidae (Shin et al. 2021). To go a step further, we attempted the functional characterization after heterologous expression of 52 of these GH5\_2 proteins derived from seven species. We show that in Cerambycidae, GH5\_2 enzymes clustered in five highly supported clades. Each clade presented different substrate specificity according to our functional analyses, including activity on abundant hemicellulose polysaccharides like xyloglucan, xylan, and mannans. Our data indicate that, at least in longhorned beetles, GH5\_2 enzymes are not monospecific. Reconstruction of the ancestral state indicated that the horizontally acquired 'original' enzyme, an endo- $\beta$ -1,4-glucanase, acted mainly on amorphous cellulose with some promiscuous catalytic activities on other PCW polysaccharides. Recent gene duplications at the species level resulted in some cases in novel enzyme activity, such as the ability to perform transglycosylation. Altogether, our data strongly indicate that the function of GH5\_2 enzymes cannot be restricted to the breakdown of the sole polysaccharide cellulose. We propose that GH5\_2 enzymes played a central role in the evolution of PCW breakdown in the beetle family Cerambycidae by allowing the larvae of most species to digest abundant hemicellulose polysaccharides, thus contributing to their ability to adapt to a PCW-rich diet, and likely playing an important role in driving their radiation.

## Results

### Ancient Gene Duplications Expanded the Substrate Specificity of Cerambycid GH5\_2

We annotated 83 GH5\_2 sequences corresponding to 17 species of Cerambycidae we analyzed in a recent transcriptome study (Shin et al. 2021). In addition, we collected 22 sequences from public databases. Finally, we extracted eight sequences from preliminary genome data of the Spondylidinae *Tetropium fuscum*. In total, we analyzed the phylogenetic relationships of 113 GH5\_2 sequences derived from 25 species distributed in five subfamilies of Cerambycidae (supplementary

table S1, Supplementary Material online). Cerambycid GH5\_2 clustered in five distinct, highly supported clades (clades I–V) according to our maximum likelihood (ML) analysis (fig. 1, supplementary data S1, Supplementary Material online). At least one GH5\_2 sequence of most species was found in each of the five clades, indicating some degrees of orthology. Exceptions exist, especially in clades I, III, IV, and V, in which species-specific gene duplications occurred (fig. 1). In addition, clade V only harbors sequences from species of Lamiinae and Spondylidinae, except for RBI3 from *Rhagium bifasciatum* (Lepturinae). Altogether, these results indicate that the common ancestor of the extant species of longhorned beetles analyzed here possessed at least four, and as many as five, GH5\_2-encoding genes.

We then asked whether the phylogenetic partition of the cerambycid GH5\_2 proteins we observed corresponded to differences in the enzyme activity of each group of paralogs. To address this question, we selected nine species representative of the five subfamilies of Cerambycidae whose species were found to harbor GH5\_2-encoding genes, and we systematically attempted to heterologously express each of their GH5\_2 proteins. From the initial 52 GH5\_2 proteins, eight did not express at all in Sf9 insect cells (fig. 1). An extra six expressed successfully, but these were not enzymatically active on any of the substrates we tested (fig. 1). In total, 38 proteins were successfully expressed in Sf9 cells and enzymatically active on at least one of the substrates we tested (fig. 2, supplementary figs. S1–S7, Supplementary Material online). We were able to classify the enzymes into five categories according to the type of substrates they were able to break down.

The ones that were exclusively active on xylan poly- and oligosaccharides were restricted to clade I (figs. 1 and 2, supplementary figs. S1 and S7, Supplementary Material online). The two previously characterized xylanases from *Anoplophora glabripennis* AGL1 (McKenna et al. 2016) and *Apriona japonica* AJA1 (Pauchet et al. 2014, 2020) also clustered in clade I (fig. 1). The pattern of breakdown products observed on thin layer chromatography (TLC) indicated that this group of GH5\_2 enzymes were endo- $\beta$ -1,4-xylanases (fig. 2, supplementary figs. S1 and S7, Supplementary Material online).

Enzymes only active on xyloglucan with very limited activity on cellooligosaccharides clustered in clade II (fig. 1). According to the pattern of reaction products observed on TLC (fig. 2, supplementary figs. S2 and S7, Supplementary Material online), which is composed of large oligosaccharides, these enzymes acted as xyloglucan-specific endo- $\beta$ -1,4-glucanases. The previously characterized AGL2 from *A. glabripennis* was found to also produce breakdown products when carboxymethyl cellulose (CMC) was used as a substrate in addition to being active on xyloglucan (McKenna et al. 2016). A similar situation was also observed for AJA3 from *A. japonica* (supplementary fig. S2, Supplementary Material online).

A group of enzymes clustering in clade III (fig. 1) was found to be highly active on CMC, regenerated amorphous

cellulose (RAC), and completely broke down cellooligosaccharides (cello-tetraose and longer) (fig. 2, supplementary figs. S3 and S7, Supplementary Material online). In addition, some activity on glucomannan, but not on galactomannan, was observed for these GH5\_2 enzymes (fig. 2, supplementary figs. S3, S6, and S7, Supplementary Material online). We named this clade ‘cellulase’ because, compared with the enzymes clustering in clades IV and V, these enzymes showed strong activity on cellooligosaccharides (see below).

Enzymes mostly active on glucomannan, but also showing some degree of activity on CMC and cellooligosaccharides, clustered together in clade IV (fig. 1). These endo- $\beta$ -1,4-mannanases can be distinguished from those clustering in clade V because the latter were active not only on glucomannan but also on galactomannan (fig. 1). Only two enzymes (AAE5 and RBIC3) located in clade IV were also active on galactomannan (fig. 2, supplementary figs. S4–S7, Supplementary Material online).

In summary, we observed a correlation between the phylogenetic clustering of cerambycid GH5\_2 and their substrate specificity. In other words, ancient gene duplication events broadened the spectrum of PCW-associated polysaccharides able to be broken down by this family of enzymes. The most ancient gene duplication, which split enzymes from clade I to all the other GH5\_2 enzymes, allowed the clade I enzymes to use a backbone made of pentose sugars as a substrate; in contrast, all the enzymes from clades II, III, IV, and V use polysaccharides having a backbone made of hexose sugars as substrates. The second most ancient gene duplication event, which split clade II to clades III + IV + V, gave rise to enzymes able to use xyloglucan as a substrate, a polysaccharide made of a ‘cellulosic’ backbone ( $\beta$ -1,4-linked glucose residues) with branched xylose residues. Finally, the final rounds of ancient duplications produce enzymes acting preferentially either on amorphous cellulose (clade III) or on glucomannan (clade IV) or galactomannan (clade V).

### The Ancestral Cerambycid GH5\_2 Acquired Through HGT was a Catalytically Promiscuous Cellulase

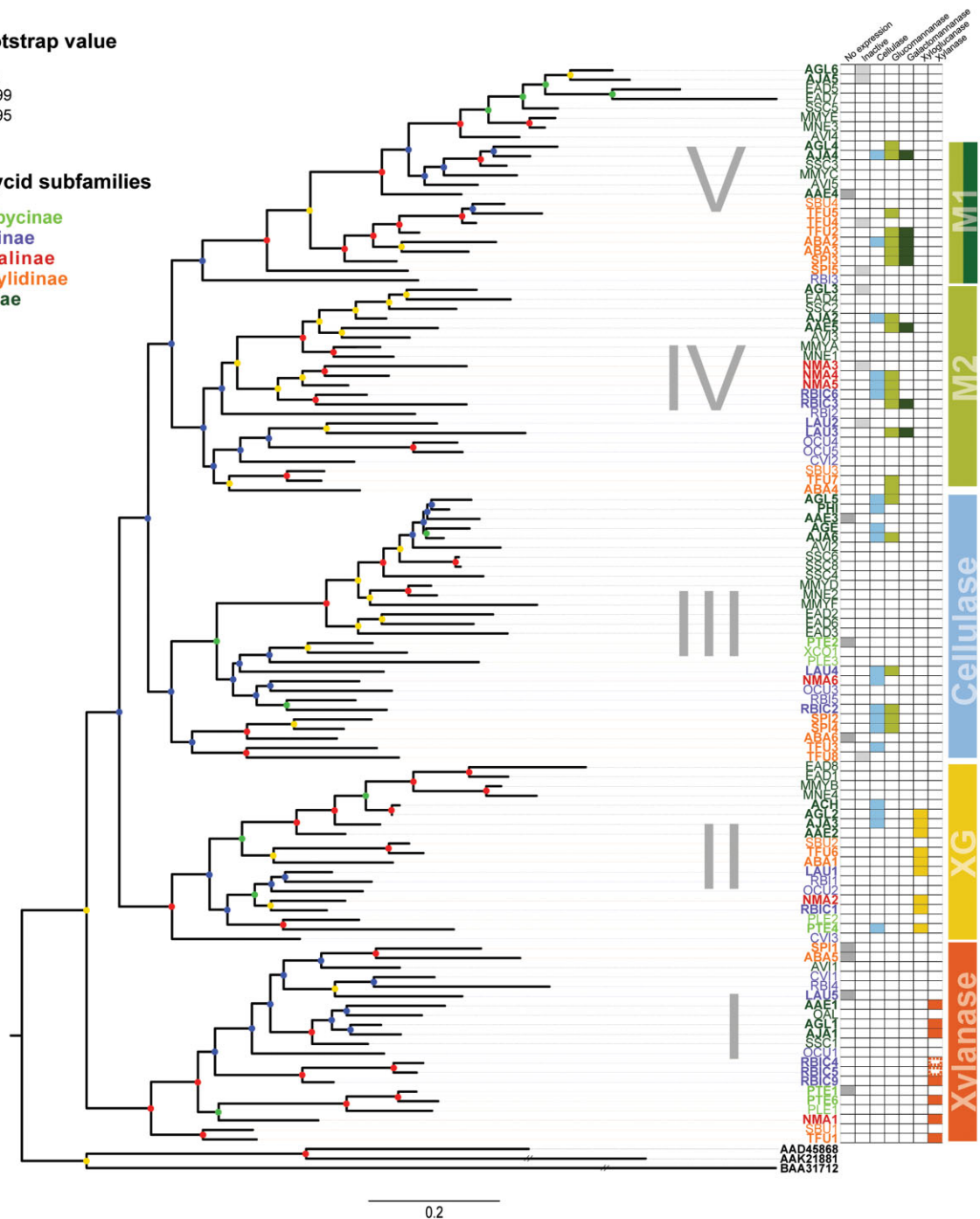
Taking into account the broad substrate specificity observed in GH5\_2 paralogs in Cerambycidae, we wondered what the function of the ancestral enzyme could have been. To address this question we aimed to resurrect the putative cerambycid ancestral GH5\_2 enzyme based on the extant enzymes found in today’s longhorned beetles. To achieve this, we used an ancestral-state reconstruction (ASR) approach based on ML (supplementary fig. S8 and data S2, Supplementary Material online). We managed to reconstruct a sequence of an ancestral GH5\_2 protein for the most ancient node (node A; fig. 3). The resulting sequence was codon-optimized, and the corresponding product was successfully expressed in insect Sf9 cells. The ancient enzyme was strongly active on polysaccharides mimicking amorphous cellulose (CMC and RAC), as well as on

## Ultrabootstrap value

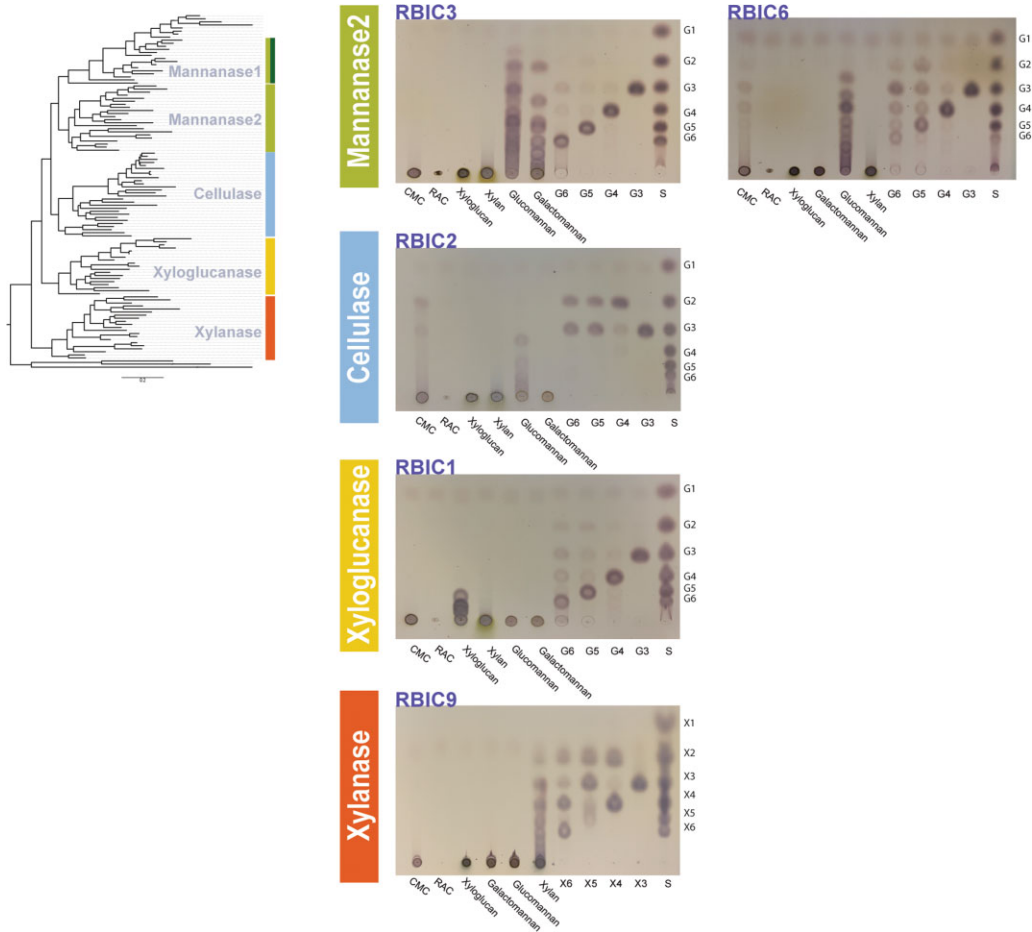
- = 100
- 96 ~ 99
- 90 ~ 95
- < 90

## Cerambycid subfamilies

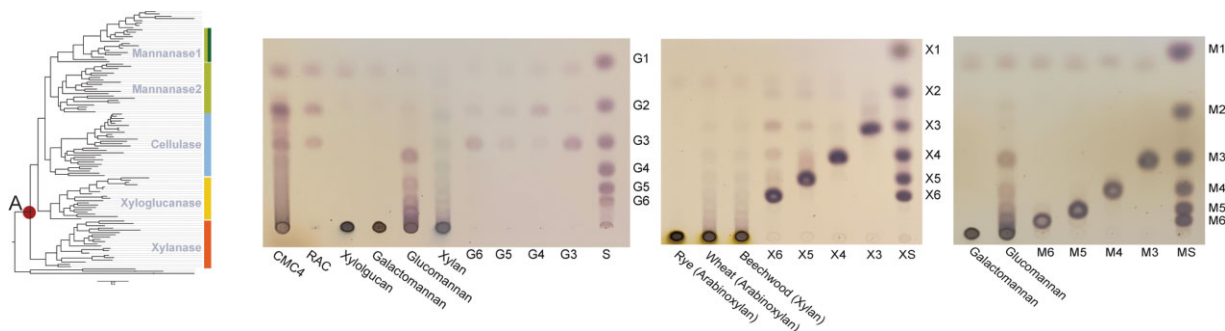
- Cerambycinae
- Lepturinae
- Necydalinae
- Spondylidinae
- Lamiinae



**Fig. 1.** Phylogenetic relationships of GH5\_2 proteins from longhorned beetles and their corresponding enzymatic activity. We performed a ML analysis in IQ-TREE with 1000 ultrafast-bootstrap replicates. 113 GH5\_2 amino acid sequences derived from 25 species of cerambycid beetles, together with two GH5\_2 derived from nematodes (AAD45868.1 and AAK21881.1) and one from bacteria (BAA31712.1) used as an outgroup, were aligned using MAFFT. The best-fit substitution model, determined using ModelFinder was the Whelan and Goldman (WAG); this model incorporated a discrete gamma distribution (shape parameter = 4) to model evolutionary rate differences among sites (+G) and contained a proportion of invariable sites (+I). Support values for each node are indicated by discs of different colors: red (equal to 100); yellow (96–99); green (90–95); and blue (below 90). We use roman numerals for individual clades according to the substrate specificity of the corresponding enzymes: clade I (xylanase), clade II (xyloglucanase), clade III (mainly cellulase), IV, and V (mainly mannanase). Individual sequence names, which contain the abbreviated species name and a number (supplementary table S1, Supplementary Material online), were labeled with different colors according to the corresponding subfamilies (light green: Cerambycinae; purple: Lepturinae; red: Necydalinae; orange: Spondylidinae; and dark green: Lamiinae). We attempted the heterologous expression of 46 GH5\_2 in Sf9 cells. Those for which no expression was obtained are marked with a dark gray square in the column ‘No expression.’ Those that were successfully expressed, but for which no enzymatic activity was detected on the substrates tested, are marked with a light gray square in the column ‘inactive.’ Enzymes active on cellulose poly- and oligosaccharides are marked with a light blue square in the column ‘cellulase.’ Those active on glucomannan are marked in light green in the column ‘glucomannanase.’ Those active on galactomannan are marked with a dark green square in the column ‘galactomannanase.’ Those active on xyloglucan are marked with a yellow square in the column ‘xyloglucanase.’ Those active on xylan poly- and oligosaccharides are marked in red in the column ‘xylanase.’



**FIG. 2.** Functional characterization of GH5\_2 enzymes from *Rhamnusium bicolor*. We show the results of enzyme assays obtained from five recombinant GH5\_2 from *R. bicolor*. Recombinant enzymes were incubated with six polysaccharides usually found in PCWs including CMC, RAC, xyloglucan, xylan, glucomannan, and galactomannan. In addition, several oligosaccharides were also tested as substrates: cellohexaose (G6), cellopentaose (G5), cellotetraose (G4), and cellotriose (G3), as well as xylohexaose (X6), xylopentaose (X5), xylotetraose (X4), and xylotriose (X3). Products generated by the individual enzymes were developed on TLC plates. Various mixtures of cellulooligosaccharides (glucose to cellohexaose) were used as standards for all *R. bicolor* GH5\_2 except RBIC9. Instead, a mixture of xylooligosaccharides (xylose to xylohexaose) was used as a standard for RBIC9. The results of enzyme assays performed with recombinant GH5\_2 of other beetle species can be found in [supplementary figs. S1–S7, Supplementary Material](#) online.



**FIG. 3.** Functional characterization of the resurrected ancestral GH5\_2 enzyme of Cerambycidae. An ancestral-state reconstruction, based on a maximum likelihood approach, was performed in MEGA using a codon-based alignment of 78 GH5\_2 sequences. The sequence of an ancestral reconstructed GH5\_2 gene (corresponding to the node marked with 'A') was codon-optimized, synthesized, and expressed in sf9 cells. The activity of the resurrected ancestral GH5\_2 enzyme was tested against six polysaccharides including CMC, RAC, xyloglucan, galactomannan, glucomannan, and xylan. Two extra xylan polysaccharides—arabinoxylan from wheat and rye—were also tested. In addition, cello- (cellotriose to cellohexaose), xylo- (xylotriose to xylohexaose), and manno-oligosaccharides (mannotriose to mannohexaose) were also used as substrates. Products were developed on TLC plates using appropriate standards (glucose to cellohexaose [G1–G6] xylose to xylohexaose [X1–X6] or mannose to manno hexaose [M1–M6]).

cellooligosaccharides (cellotetraose to cellohexaose) (fig. 3). In addition, breakdown products were also observed on TLC when glucomannan was used as a substrate (fig. 3). Interestingly, this resurrected enzyme possessed the ability to break down to some extent xylan and xylooligosaccharides (fig. 3). In contrast, no breakdown products were observed when manno-oligosaccharides were used as substrates (fig. 3). Altogether, we conclude that the ancestral GH5\_2 enzyme in longhorned beetles, which was acquired from a bacterial donor through HGT, was an endo- $\beta$ -1,4-glucanase enzyme; although mainly active on amorphous cellulose, the enzyme possesses some degree of catalytic promiscuity and could break down other polysaccharides of the PCW, such as glucomannan and xylan.

### Recent Gene Duplications Broadened the Enzymatic Capabilities of Cerambycid GH5\_2 Even More

Looking at the phylogenetic tree of cerambycid GH5\_2 (fig. 1), we observed a few examples of species-specific gene duplications. We wondered what the outcome of such recent duplication events was.

In clade IV, three recently duplicated GH5\_2 proteins from the species of Necydalinae *Necydalis major* were present (fig. 1). Although NMA3 was found to be inactive on all the substrates tested, NMA4 and NMA5 were enzymatically active on the same substrates, mostly glucomannan and, to a lesser extent, cellopentaose and cellohexaose. Both enzymes produced very similar breakdown products on TLC (supplementary fig. S4, Supplementary Material online). This gene duplication event represents a good example of increased gene dosage. Increase in gene dosage was also observed for other recent duplicates, such as SPI2 and SPI4 (*Saphanus piceus*; Spondylidinae) clustering in clade III (fig. 1, supplementary fig. S3, Supplementary Material online).

Also in clade IV, the duplicated RBIC3 and RBIC6 (*Rhamnusium bicolor*; Lepturinae) were both found to act on glucomannan, which is common for other enzymes clustering in this clade (figs. 1 and 2). However, RBIC3 evolved the ability to also use galactomannan as a substrate (fig. 2), which is unusual for enzymes clustering in clade IV but common for those clustering in clade V (fig. 1). This observation correlated well with the lack of GH5\_2 from *R. bicolor* clustering in clade V.

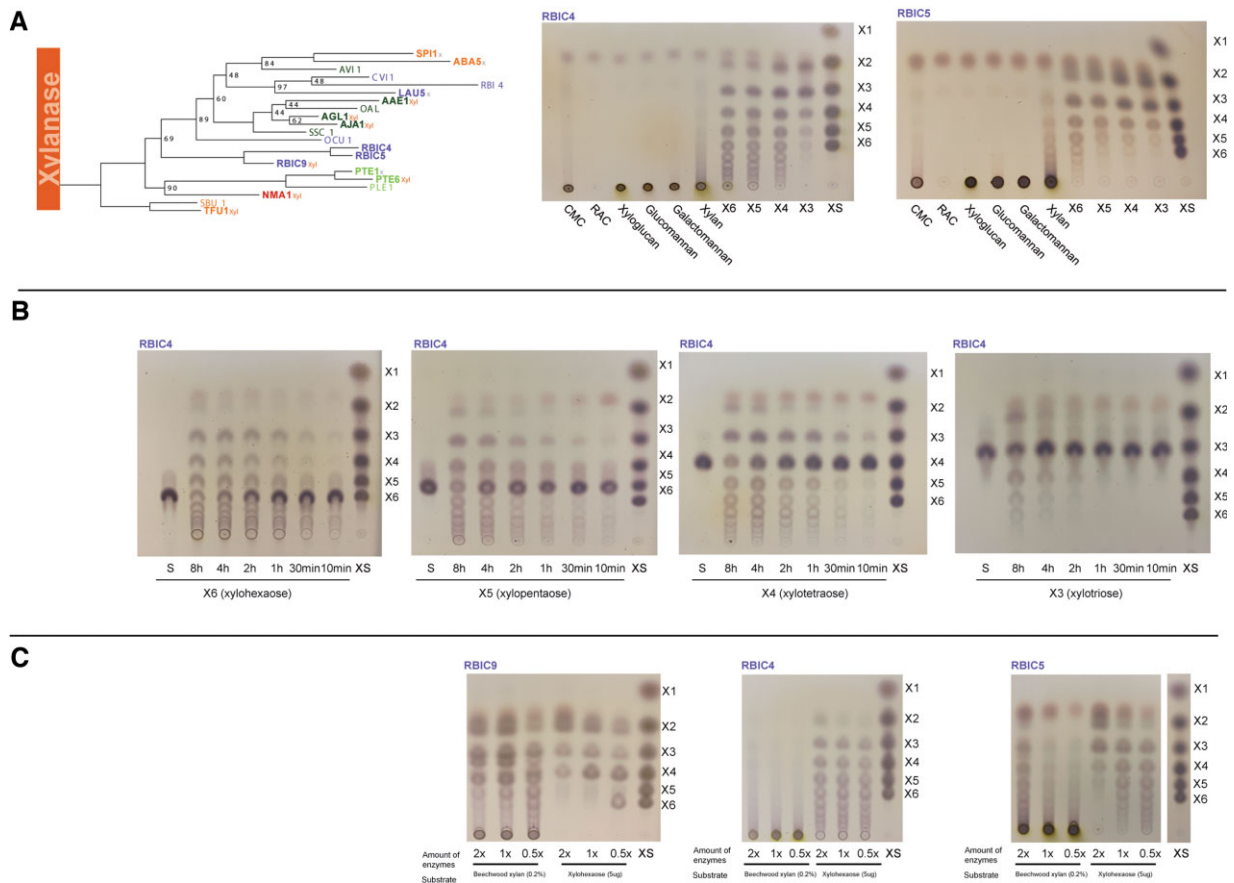
We observed the most striking case of subfunctionalization for three duplicates derived from *R. bicolor* (RBIC4, RBIC5, and RBIC9) and clustering into clade I (figs. 1 and 4A). A first duplication event gave rise to RBIC9 and RBIC4 + RBIC5, which later further duplicated. RBIC9 behaved like other endo- $\beta$ -1,4-xylanases from clade I, breaking down xylan oligo- and polysaccharides into smaller products (fig. 2). Unexpectedly, in our standardized assay conditions, RBIC4 was barely active on xylan polysaccharides but produced a 'ladder-like' pattern of oligosaccharides ranging from xylobiose to at least xylohexaose, according to our TLC, when incubated with xylohexaose, and, to a lesser extent,

with xylotriose (fig. 4A). This observed pattern indicates that RBIC4 may have an increased ability to perform transglycosylation. In end-point determination experiments, RBIC5 generated a slightly different pattern of products compared with RBIC4. First, breakdown products were observed when RBIC5 was incubated with a xylan polysaccharide. Second, the 'ladder-like' pattern of end-products obtained using xylohexaose to xylohexaose was less obvious than the patterns produced by RBIC4 (fig. 4A).

We went deeper in the functional characterization of these three *R. bicolor* enzymes (fig. 4). First, we performed a time-course experiment with RBIC4 incubated with single xylooligosaccharides ranging from xylohexaose down to xylotriose (fig. 4B). We observed the accumulation of oligomers—larger than the substrate provided—as early as 10 min after the incubation started. The accumulation of hydrolysis products (oligomers smaller than the substrate provided) followed soon after. The reactions seemed to reach equilibrium after 4–8 h, as the pattern of products stopped changing (fig. 4B). Then, we varied the amount of enzyme (for RBIC4, RBIC5, and RBIC9) but kept a fixed concentration of substrate, either xylan or xylohexaose (fig. 4C). The release of hydrolysis products improved with increasing amounts of RBIC9 in the reaction. We observed no variation of the pattern of products with more RBIC4 in the reaction. However, increasing the amount of RBIC5 seemed to push the reaction towards hydrolysis, when incubated either with a xylan polysaccharide or with xylohexaose. The lesser the amount of RBIC5 in the reaction, the higher the amount of transglycosylation products observed on the TLCs, indicating that RBIC5 was more sensitive than RBIC4 to changes in the ratio of enzyme to substrate (fig. 4C).

### Can Positive Selection Explain the Transglycosylation Ability of *rbic4* and *rbic5*?

The discovery of enzymes, which seem to favor transglycosylation rather than hydrolysis, made us wonder whether amino acids under positive selection might provide insights on how this transition happened. To find out, we used a branch-site model approach and analyzed patterns of positive selection based on the protein-coding sequences of 101 cerambycid GH5\_2. From the resulting ML tree, patterns of positive selection could be detected on 8 out of 11 tested branches (supplementary table S3, fig. S9, and data S3, Supplementary Material online). Interestingly, two of these branches corresponded to the two duplication events that gave birth to RBIC9 (the xylan hydrolase) and to RBIC4 and RBIC5 (the xylan transglycosidases) (supplementary fig. S9, Supplementary Material online). Bayes empirical Bayes (BEB) analyses identified codon positions having signs of positive selection on these two branches (supplementary table S3, Supplementary Material online). According to our analyses, only three amino acid positions were found to be



**Fig. 4.** Detailed functional characterization of three xylan-active duplicates of *Rhamnusium bicolor*. (A) We assessed the function of RBIC4 and RBIC5, two duplicates clustering together with RBIC9 that was found to be an endo- $\beta$ -1,4-xylanase as exemplified in figure 2. Recombinant RBIC4 and RBIC5 were incubated with six polysaccharides including CMC, RAC, xyloglucan, galactomannan, glucomannan, and xylan, as well as xylooligosaccharides, and the resulting products were developed on TLC plates. (B) Time-course experiments were performed with RBIC4 and xylohexaose, xylopentaose, xylotetraose, and xylotriose as substrates. Enzyme assays were incubated between 10 and 8 h, and the resulting products were developed on TLC plates. S: substrate incubated without enzyme for 8 h. (C) Increasing amounts of recombinant RBIC9, RBIC4, and RBIC5 (1X is the equivalent of 0.65  $\mu$ l of anti-V5 agarose beads) were incubated with fixed amounts of either beechwood xylan (0.2% final concentration) or xylohexaose (0.25  $\mu$ g/ $\mu$ l). Enzyme assays were incubated for 16 h, and the products were developed on TLC plates. Xylose (X1) to xylohexaose (X6) were used as standards.

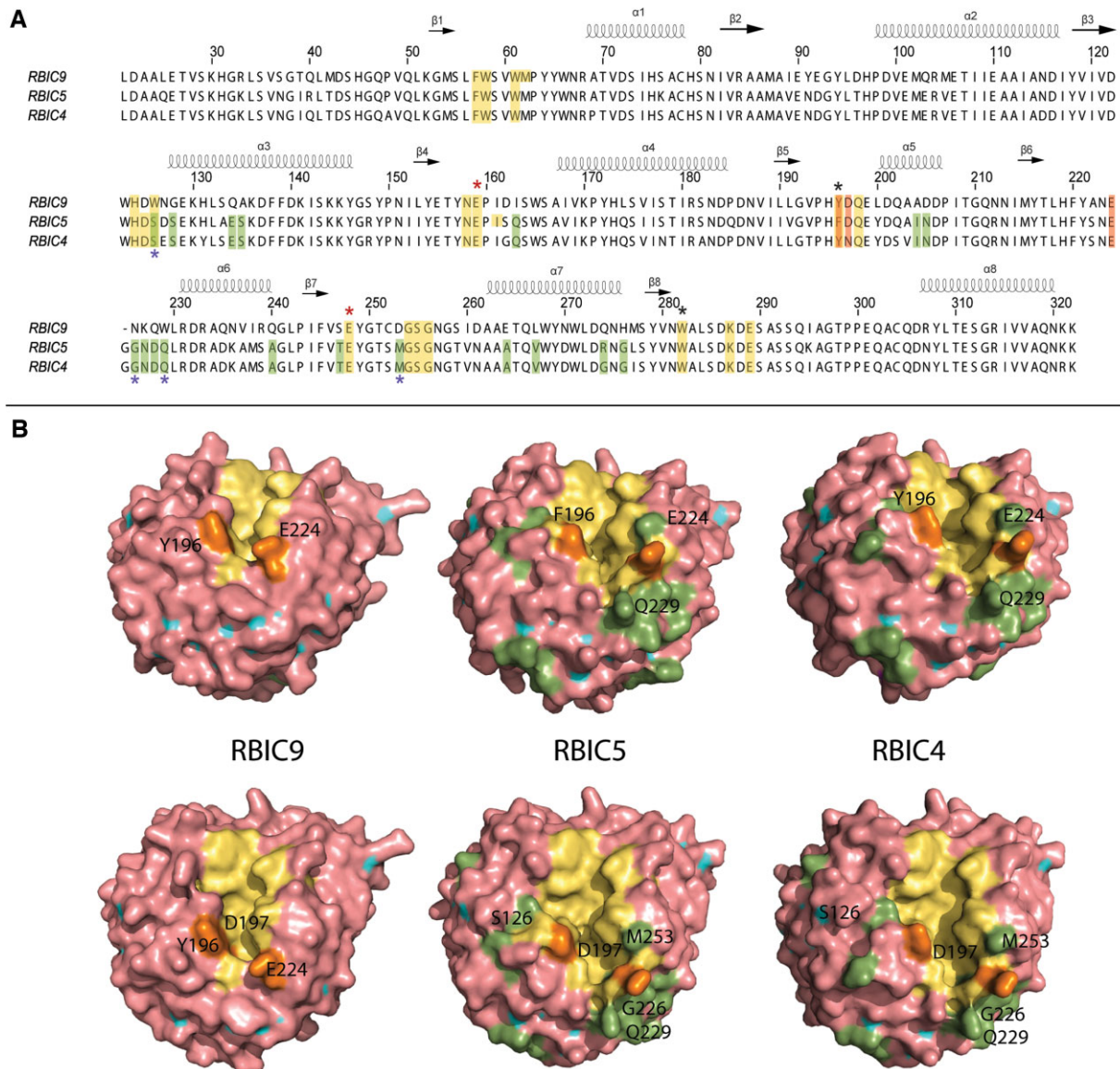
under positive selection on branch R9 (supplementary fig. S9, Supplementary Material online), the one leading to the first split between RBIC9 and RBIC4 + RBIC5. We then plotted the positions of the corresponding amino acids: (1) on a sequence alignment (fig. 5A); and (2) on reconstructed three-dimensional models of RBIC9, RBIC4, and RBIC5 (fig. 5B). These amino acids were either part of the active site or located close to it (fig. 5B). An extra 18 amino acid positions were detected to be under positive selection on branch gt (supplementary fig. S9, Supplementary Material online), the one leading to the split between RBIC4 and RBIC5. Most of these positions were located at the surface of the protein relatively far away from the catalytic pocket (fig. 5B). However, a few were found within or nearby the entrance of the catalytic pocket (fig. 5B). Accordingly, we suggest that these amino acids should be the first to be assessed for their potential role in the transition between hydrolysis and transglycosylation through mutagenesis experiments.

## Discussion

### Three Level of Functional Novelty Shaped the Evolution of GH5\_2 in Cerambycidae

The genomes of phytophagous beetles are called the Phytophaga—which encompasses the two superfamilies: (1) Chrysomeloidea (leaf beetles and longhorned beetles); and (2) Curculionoidea (weevils and bark beetles)—encode an arsenal of PCWDEs that helps the insects digest their plant-based diet. Some of these families of enzymes are widely distributed among species of Phytophaga, such as GH45 cellulases and GH28 polygalacturonases (Pauchet et al. 2010; Kirsch et al. 2012, 2014; Busch et al. 2019; McKenna et al. 2019). Other enzyme families are more restricted in their distribution, which is the case for GH5\_2. To date, this enzyme family has been restricted to the longhorned beetle family Cerambycidae (McKenna et al. 2019; Shin et al. 2021).

The aim of our study was to investigate the events that shaped the evolutionary history of GH5\_2 enzymes in this

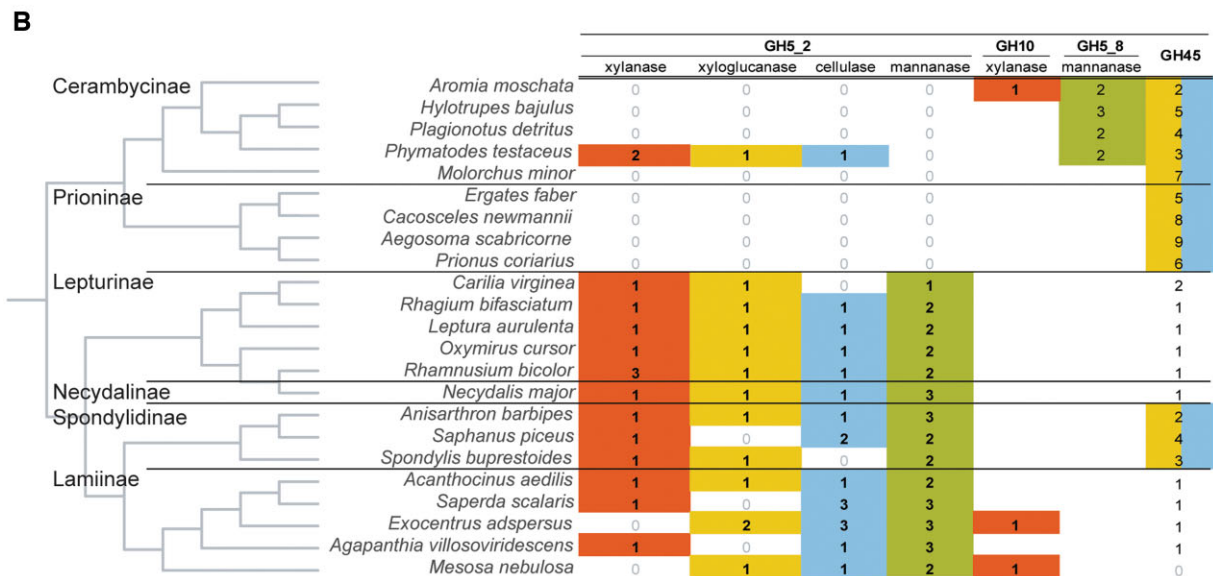
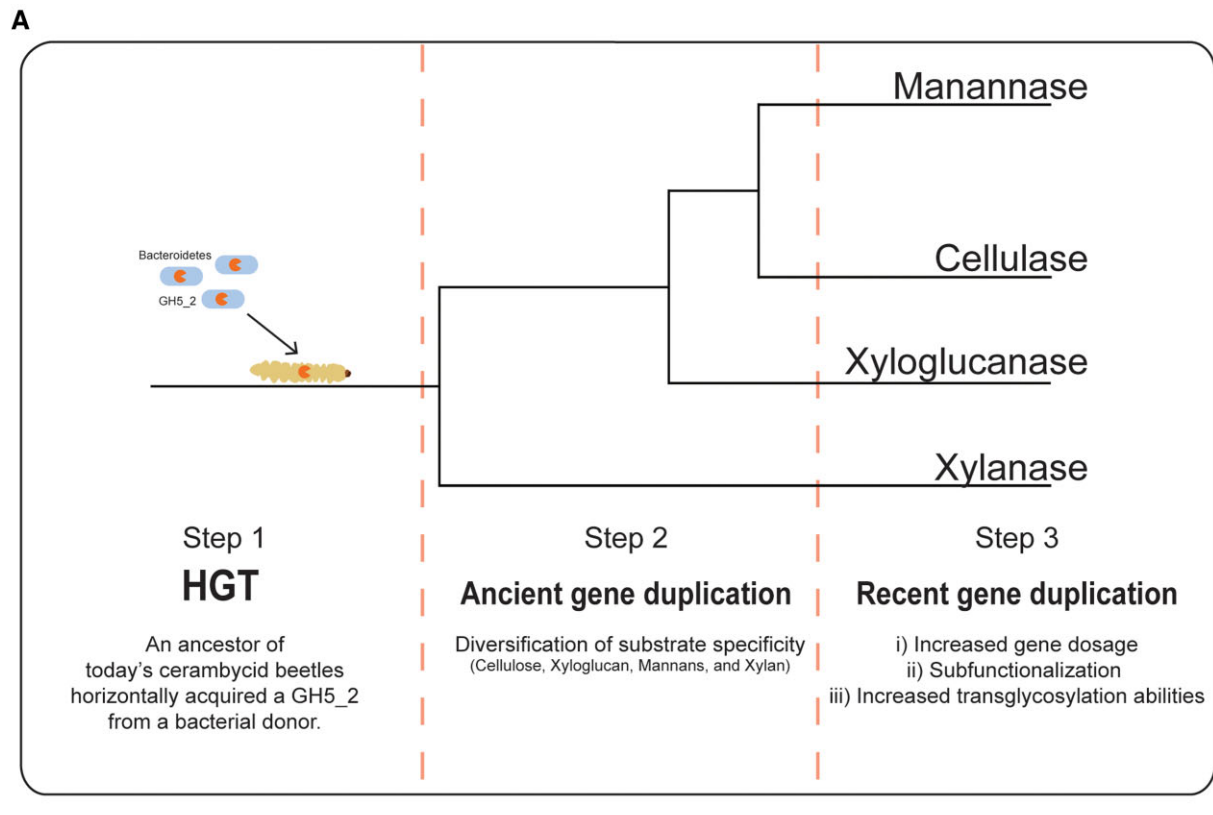


**Fig. 5.** Localization of amino acid positions under positive selection leading to the *Rhammusium bicolor* transglycosidases. (A) The amino acid sequences of RBIC9, RBIC5, and RBIC4 were aligned using MAFFT. Protein structural information, including the position of alpha-helices ( $\alpha_1$ – $\alpha_8$ ) and of beta-sheets ( $\beta_1$ – $\beta_8$ ) are provided above the sequence alignment. Black asterisks indicate the conserved substrate binding sites according to published structures of GH5\_2 proteins. Red asterisks correspond to the two conserved catalytic glutamates. Additional residues that are part of the active site, estimated by our homology modeling analysis using MOE-Site Finder, are shaded in yellow. Amino acid positions with signal of positive selection estimated by a branch-site model and BEB analyses for branch R9 (supplementary fig. S9, Supplementary Material online) are shaded in orange and in green for branch gt (supplementary fig. S9, Supplementary Material online). Purple asterisks represent putative additional binding sites from structural comparison that was found to be under positive selection. (B) The above-mentioned amino acid positions—using the same color code—are visualized on three-dimensional protein structures of RBIC9, RBIC4, and RBIC5 generated using the SWISS-model.

clade of beetles. From our results, a pattern emerged according to which functional novelties appeared at three different levels (fig. 6A). First, the acquisition through HGT of a GH5\_2 cellulase from a bacterial donor in the common ancestor of today's extant cerambycid species represented a major shortcut in the evolution of novel digestive abilities. Second, by broadening the number of PCW polysaccharides that could be used as substrates by these enzymes, ancestral gene duplication expanded the digestive capacities of these beetles. This step was likely

promoted by the promiscuous characteristic of the ancestral, horizontally acquired, enzyme. Third, recent gene duplications at the genus/species level either led to an increased gene dosage or expanded the digestive capabilities of the corresponding beetles (fig. 6A). One of our more striking discoveries was two GH5\_2 paralogs—for which the corresponding sequences clustered within the xylanase clade—that seemed to prefer transglycosylation over hydrolysis, in contrast to other functionally characterized enzymes from the same clade.





**FIG. 6.** Summary of the evolutionary history of GH5\_2 proteins in Cerambycidae according to our data. (A) We propose three major steps that contributed to the evolution of GH5\_2 enzymes in extant species of Cerambycidae. Step 1: A GH5\_2 cellulase was horizontally transferred from a bacterial donor to the most recent common ancestor of the Cerambycidae. Step 2: At the beginning of the radiation that gave rise to modern Cerambycidae gene duplications occurred with an increase of substrate specificity, with paralogs evolving the ability to break down most polysaccharides composing hemicellulose and cellulose. Step 3: Genus/species-specific gene duplications allowed for either increased gene dosage in some cases or subfunctionalization, or for the evolution of enzymes predominantly performing transglycosylation instead of hydrolysis. (B) Schematized species tree according to our previous transcriptome analysis (Shin et al. 2021). The number of enzymes from each category/family is indicated in the table on the left hand side of the figure. The absence of given categories of GH5\_2 at the species level can be compensated for by the presence of other GH families. For example, the absence of a GH5\_2 xylanase in *Exocentrus* and *Mesosa* can be compensated for by the presence of a GH10 xylanase found in the larval gut transcriptomes of the corresponding species.

**Relationship Between GH5\_2 and Other GH Families**  
Our results illustrated that cerambycid GH5\_2 enzymes can be categorized according to their substrate specificity

(fig. 1). The extent of our analyses—combining phylogenetic relationships to functional characterization of the corresponding enzymes—reached such a level that almost

any newly discovered cerambycid GH5\_2 in the future could be attributed a function according to its phylogenetic relationships. Although some of the beetle species we analyzed here harbor at least one GH5\_2 enzyme per category, some species lack GH5\_2 representatives in some of these categories, especially for xylanase, xyloglucan-specific endo- $\beta$ -1,4-glucanase, and cellulase (fig. 6B). In this context, we argue that other families/sub-families of GHs could compensate for such a lack of function. Both lamiine species, *Exocentrus adpersus* and *Mesosa nebulosa*, lack a GH5\_2 xylanase, but we noted the presence of a sequence corresponding to a GH10 protein, a well-characterized family of xylanases, in their respective transcriptome (fig. 6B). Similarly, *Phymatodes testaceus* (Cerambycinae) lack GH5\_2 mannanases completely, but we found several copies of putative GH5\_8 mannanases in the transcriptome of this species. The presence of GH5\_8 seems to be a signature of Cerambycinae transcriptomes, apart from the basal *Molorchus minor* (fig. 6B). In species of longhorned beetles completely lacking GH5\_2, such as *M. minor* (Cerambycinae) and the four species of Prioninae we analyzed, we observe a striking increase in the copy number of GH45 proteins compared with species harboring GH5\_2 paralogs (fig. 6). These enzymes have been extensively studied in the past and most of them are endo- $\beta$ -1,4-glucanases mostly acting on amorphous cellulose ([www.cazy.org](http://www.cazy.org)). Recent studies on species of leaf beetles, which are members of the same beetle superfamily as longhorned beetles (the Chrysomeloidea), demonstrated that some GH45 paralogs diversified their substrate specificity to not only break down amorphous cellulose, but also xyloglucan and glucomannan (Busch et al. 2018, 2019). We hypothesize that in species of longhorned beetles lacking GH5\_2, GH45 paralogs could take over the degradation of amorphous cellulose, xyloglucan and mannans.

We wondered what could have fueled such a dynamic evolutionary scenario. The first possibility could be the horizontal acquisition of an enzyme being already more catalytically efficient than its GH5\_2 counterpart. In this context, the latter enzyme, because it would be less needed, would be progressively selected against and eventually lost. Second, environmental factors such as the quality of the food source may also play a role. Most Cerambycidae are xylophagous, but their feeding habits can differ drastically: some feed on living tissue of a tree, such as the cambium, some feed on dead and/or dry wood or even on highly decayed wood (Haack 2017). In this context, the range of PCWDEs needed to digest efficiently in these different types of diets may vary greatly and influence which enzymes are selected for or against.

### Why Maintain a Transglycosidase Activity?

Taking into account that the main function of digestive enzymes is to break down macromolecules to smaller ones in order to facilitate their absorption or to get access to other nutrients, finding an enzyme that favors transglycosylation

over hydrolysis is puzzling. A growing body of evidence suggests that the gut microbiota may have a significant impact on the health of many insects in general and longhorned beetles in particular (Schloss et al. 2006; Kim et al. 2017). In addition, such commensal bacteria have also been suggested to play a role in adaptation to the host plant (Ge et al. 2021). Oligosaccharides—in particular, xylooligosaccharides—have long been known to act as prebiotics, helping to maintain a healthy gut community and a strong immune system in animals (Vazquez et al. 2000). In this context, we suggest that RBIC4, and to a lesser extent RBIC5, may stimulate the gut microbiota in larvae of *Rhamnusium bicolor* by supplementing it with a variety of xylooligosaccharides of different sizes.

### Novel Enzymes for Biotechnology

Carbohydrates are ubiquitous in nature and play crucial roles in a number of biological processes. In this context, there is growing interest in the in vitro synthesis of well-defined carbohydrate compounds, not only for use in basic research but also for the preparation of commercially valuable products. For example, some oligosaccharides and their derivatives can be used in the food industry as prebiotics, or have multiple therapeutic and cosmetic uses (Vazquez et al. 2000). Strategies for the synthesis of oligosaccharides include chemical methods, which are tedious with low yields (Santibanez et al. 2021), and the use of enzymes. Enzyme-driven synthesis of oligosaccharides can be achieved by glycosyltransferases (GTs). However, the use of GTs on industrial scale is hampered by the fact that these enzymes are sparse, unstable in solution, and require difficult-to-produce nucleotide sugars as donor substrates (Bissaro et al. 2015). GHs represent an interesting alternative to GTs for the synthesis of oligosaccharides. Like most of the GH5\_2 enzymes we tested here, GHs usually perform the hydrolysis of polysaccharides into shorter oligomers, but in given conditions—often dependant of pH and temperature—they can also perform transglycosylation, which is the reverse reaction, and synthesize long oligomers using shorter ones as a substrate (Bissaro et al. 2015). The discovery of an enzyme, such as RBIC4, which seems to naturally favor transglycosylation over hydrolysis in our standardized assay conditions, exemplifies why our approach is powerful, namely, it allows to systematically test the enzymatic properties of all the members of a given enzyme family in several species. Taking into account the discrepancy between the number of GH sequences available in public databases and how many of them have actually been functionally characterized, we anticipate that many more catalytically relevant enzymes will be found.

### Concluding Remarks

Here, we exhaustively analyze the evolution of a family of glycoside hydrolases in an understudied clade of phytophagous insects. Using reverse genetics, that is, CRISPR-Cas9 knockouts of individual GH5\_2 enzymes, would have made our study even more complete. However, such molecular tools so far do not exist for this group of insects for three reasons: they are very

difficult to rear in the laboratory, larvae of most species are naturally concealed inside branches or trunks of trees, and some species may take up to several years to develop. Yet because we are fascinated by how substrate specificity evolved in cerambycid GH5\_2 enzymes, including the transition to transglycosylation, we will attempt in the future to crystalize and determine the three-dimensional structure of representative GH5\_2 enzymes of each relevant clade. Altogether, given that GH5\_2 enzymes are distributed only in species of longhorned beetles and are absent in other group of Phytophaga beetles, such as leaf beetles and weevils, we argue that this enzyme family represents a crucial step in the evolution and extraordinary radiation of a mostly xylophagous clade of insects.

## Materials and Methods

### Phylogenetic Analysis of Cerambycid GH5\_2

We recovered all the GH5\_2 sequences annotated from our previous transcriptome analysis. In addition, we searched the nonredundant protein database at NCBI for other, previously characterized, cerambycid GH5\_2 sequences. Predicted signal peptides were determined using SignalP v5.0, and their corresponding sequences were removed. Alignments of the final 113 GH5\_2 amino acid sequences were performed using MAFFT v7.471 (Kuraku *et al.* 2013). The resulting alignments were inspected and manually adjusted when necessary. We then used IQ-TREE v2.0.3 (Nguyen *et al.* 2015) to perform ML phylogenetic analyses. The best-fit substitution model was determined within IQ-TREE using ModelFinder (Kalyaanamoorthy *et al.* 2017). Branch support was estimated by ultrafast bootstrap approximation within IQ-TREE using UFboot (Minh *et al.* 2013). Sequence alignment, tree file (Newick format), and IQ-TREE log file can be found in [supplementary data S1, Supplementary Material](#) online.

### Heterologous Expression in Insect Sf9 Cells

We used the same procedure as described elsewhere (Pauchet *et al.* 2020). Briefly, open reading frames (ORFs) of GH5\_2, excluding the stop codon, were amplified by polymerase chain reaction (PCR) using RACE-ready cDNA generated in our previous study (Shin *et al.* 2021). A Kozak sequence was added at the 5'-end of the PCR product by integrating it into the forward PCR primer. The resulting PCR products were cloned into pIB/V5-His TOPO/TA (Invitrogen, Waltham, MA, USA) in an ORF with a V5-(His)<sub>6</sub> epitope at the carboxyl-terminus, and constructs in the correct orientation were selected after colony PCR. For two constructs (RBIC5 and RBIC9), codon-optimized synthetic constructs cloned into pIB/V5-His TOPO/TA were obtained from the company GenScript (Piscataway, NJ, USA). Insect Sf9 cells (Invitrogen) were routinely cultured in SF-900 II serum-free medium (Gibco, Paisley, UK). Cells were transfected in six-well plates using FUGENE HD (Promega, Madison,

WI, USA) as the transfection reagent. After 72 h, the culture medium of transfected cells was harvested, and cell debris was removed by centrifugation. Recombinant GH5\_2 proteins were recovered by immunoprecipitation using anti-V5 agarose beads (V5-Trap, ChromoTek, Planegg-Martinsried, Germany). After immunoprecipitation, agarose beads were resuspended in 150  $\mu$ l of double-distilled water.

### Enzyme Assays

GH5\_2 proteins bound to anti-V5 agarose beads were incubated with polysaccharides usually found in PCWs. Enzyme assays (20  $\mu$ l) were set by mixing agarose beads resuspended in water (14  $\mu$ l) with a 1% solution of substrate (4  $\mu$ l) in a 20 mM citrate/phosphate buffer pH 5.0. The substrates used were CMC (Sigma-Aldrich, Saint-Louis, MO, USA), RAC, prepared as described in Busch *et al.* (2018), xyloglucan from Tamarind seeds (Megazyme, Bray, Ireland), glucomannan from konjac (Megazyme), galactomannan from carob (Megazyme), beechwood xylan (Sigma-Aldrich), arabinoxylan from rye (Megazyme), and arabinoxylan from wheat (Megazyme). In addition, oligosaccharides (all purchased from Megazyme) were also used as substrates. Enzyme assays (20  $\mu$ l) using oligosaccharides as substrates were set up as follows: GH5\_2 proteins bound to anti-V5 agarose beads (14  $\mu$ l) were mixed with a given oligosaccharide (0.5  $\mu$ l; 10  $\mu$ g/ $\mu$ l) in a 20 mM citrate/phosphate buffer pH 5.0. We tested the enzymes with celotriose to celohexose and, alternatively, with xylotriase to xylohexaose or mannotriose to mannohexaose. Enzyme assays were incubated for 16 h at 40 °C before being applied to TLC plates (silica gel 60, 20  $\times$  20 cm, Merck, Darmstadt, Germany). TLC plates were developed for a minimum of 180 min in a mobile phase composed of ethyl acetate/acetic acid/formic acid/water in the ratio 9:3:1:4, and then dried at room temperature. The hydrolysis products were subsequently revealed by soaking the plates in 0.2% (w/v) orcinol in methanol/sulfuric acid (9:1), then heated briefly until spots appeared on the plates.

### Ancestral-State Reconstruction

The amino acid sequences of 77 newly annotated GH5\_2 sequences were aligned using MAFFT v7.471 and further converted into a codon-based nucleotide alignment using PAL2NAL v14.0 (Suyama *et al.* 2006). The resulting sequence alignment was implemented in MEGA v7.0.26 using the option 'infer ancestral sequence (ML)'. The following parameters were used: the general time reversible model was selected to incorporate a discrete gamma distribution (shape parameter = 5) to model evolutionary rate differences among sites (+G) and a proportion of invariable sites (+I). MEGA returns ancestral-state reconstructed sequences automatically. Sequence alignment, the Newick file, and the output file from MEGA can be found in [supplementary data S2, Supplementary Material](#) online. The selected ancestral reconstructed sequence was used to generate a codon-optimized synthetic construct

(Genscript), which was further cloned (becoming pIB/V5-His TOPO/TA) for heterologous expression in insect Sf9 cells as described above. The enzyme activity of the obtained recombinant protein was tested as described above, and hydrolysis products were developed on TLC plates.

### Detection of Amino Acids Positions Under Positive Selection

Branch-site model tests were performed on GH5\_2 encoding genes to detect branches, and sites experiencing positive selection using the CodeML program v3.2.3 from the PAML package (Yang 1997). We performed codon-based alignment of 101 GH5\_2 encoding nucleotide sequences with two nematode and three bacterial GH5\_2 encoding genes, using MAFFT v7.471 and PAL2NAL v14.0 as described above. A ML tree was reconstructed in IQ-TREE v2.0.3. The best-fit substitution model was automatically determined within IQ-TREE using ModelFinder, and was found to be the Whelan and Goldman incorporating a FreeRate model (+R; number of categories = 5) (supplementary fig. S9, Supplementary Material online). Branch support was estimated by Ultrafast Bootstrap Approximation within IQ-TREE using UFboot. The presence of sites potentially under positive selection was analyzed on 11 different branches of the ML tree. We used the following parameters: Nssites = 2 for both the alternative and the null model, fixed\_omega = 0 for the alternative model, and fixed\_omega = 1 for the null model. We ran a likelihood ratio test with a  $\chi^2$  distribution to evaluate significant differences between the alternative and null models. *P*-values were adjusted with false discovery rates. We used BEB analysis to identify positively selected sites with posterior probability cut-off numbers of 0.99. Sequence alignment, newick file, and IQ-TREE log file can be found in supplementary data S3, Supplementary Material online.

### Homology Modeling

Homology modeling of the *R. bicolor* GH5\_2 proteins was performed using SWISS-model (Waterhouse et al. 2018). The best-fit published protein model structure was the GH5\_2 from the *Cytophaga hutchinsonii* Endoglucanase (PDB accession: 5IHS) (Zhu et al. 2016). The catalytic residues were predicted by the Site Finder module in the program Molecular Operating Environment (MOE; Chemical Computing Group, Montreal, Canada), using a geometric approach to calculate putative binding sites in a protein, starting from its tridimensional structure. Structure building and labeling of sites under positive selection were performed in PyMOL v4.5.

### Supplementary Material

Supplementary material is available at *Molecular Biology And Evolution* online.

### Acknowledgements

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch (MPI-CE, Jena) for technical support. We thank Emily Wheeler, Boston, for editorial assistance. We

also thank Roy Kirsch, Yu Okamura, David G. Heckel, and Martin Kaltenpoth (both MPI-CE, Jena) for fruitful discussions. This work was supported by the Max Planck Society and by the Deutsche Forschungsgemeinschaft (DFG PA2808/4-1 to Y.P.) and by the Invasive Species Centre (14-SIPR-FO-DOU-01 to D.D.).

### Conflict of interest

N.R.S. and Y.P. have filed a patent (EP22151474.8) related to the transglycosidase properties of RBIC4. They declare no other competing interests.

### Data Availability

All newly described GH5\_2 sequences were submitted to NCBI Genbank with accession numbers: OM585470–OM585473, OM585475–OM585494, and OM810302–OM810309. The resurrected GH5\_2 ancestral sequence was also submitted to Genbank with accession number OM810301.

The following supporting data are freely accessible on EDMOND—the Open Research Data Repository of the Max Planck Society with DOI:

- [Supplementary data S1, Supplementary Material](#) online relates to the ML analysis presented in figure 1 and contains the corresponding amino acid sequence alignment, the tree file (Newick format), and the log file from IQ-TREE.
- [Supplementary data S2, Supplementary Material](#) online relates to the ASR and contains the codon-based sequence alignment, the output of the ML-based ASR generated in MEGA, and the corresponding tree file (Newick format).
- [Supplementary data S3, Supplementary Material](#) online relates to the positive selection analysis presented in figure 5 and supplementary figure S9 and table S3, Supplementary Material online, and contains the corresponding sequence alignment, the tree file (Newick format), and the IQ-TREE log file.

Additional supporting information may be found online in the [supplementary information, Supplementary Material](#) online section at the end of the article ([supplementary tables S1–S3 and figs. S1–S9, Supplementary Material](#) online).

### References

- Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol.* **12**:1–16.
- Barras F, Bortoli-German I, Bauzan M, Rouvier J, Gey C, Heyraud A, Henrissat B. 1992. Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from *Erwinia chrysanthemi*. *FEBS Lett.* **300**:145–148.
- Bissaro B, Monsan P, Faure R, O'Donohue MJ. 2015. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem J.* **467**:17–35.

- Busch A, Danchin EGJ, Pauchet Y. 2019. Functional diversification of horizontally acquired glycoside hydrolase family 45 (GH45) proteins in phytophaga beetles. *BMC Evol Biol.* **19**:100.
- 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 Mol Biol.* **27**:633–650.
- Chang CJ, Wu CP, Lu SC, Chao AL, Ho TH, Yu SM, Chao YC. 2012. A novel exo-cellulase from White spotted Longhorn beetle (*Anoplophora malasiaca*). *Insect Biochem Mol Biol.* **42**:629–636.
- Danchin EG, Rosso MN, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A.* **107**:17651–17656.
- David LA, Alm EJ. 2011. Rapid evolutionary innovation during an archaic genetic expansion. *Nature* **469**:93–96.
- Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**:D571–D577.
- Ge SX, Shi FM, Pei JH, Hou ZH, Zong SX, Ren LL. 2021. Gut bacteria associated with *Monochamus saltuarius* (Coleoptera: Cerambycidae) and their possible roles in host plant adaptations. *Front Microbiol.* **12**:687211.
- Gilkes NR, Henrissat B, Kilburn DG, Miller RC Jr, Warren RA. 1991. Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families. *Microbiol Rev.* **55**:303–315.
- Haack RA. 2017. Feeding biology of cerambycids. In: Wang Q, editors. *Cerambycidae of the world: biology and pest management*. Boca Raton (FL): CRC Press. p. 105–132.
- Hanks LM. 1999. Influence of the larval host plant on reproductive strategies of cerambycid beetles. *Annu Rev Entomol.* **44**:483–505.
- Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon J-P, Davies G. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A.* **92**:7090–7094.
- Henrissat B, Claeysens M, Tomme P, Lemesle L, Mornon JP. 1989. Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81**:83–95.
- Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* **11**:97–108.
- Jenkins J, Lo Leggio L, Harris G, Pickersgill R. 1995. Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett.* **362**:281–285.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* **14**:587–589.
- Kim JM, Choi MY, Kim JW, Lee SA, Ahn JH, Song J, Kim SH, Weon HY. 2017. Effects of diet type, developmental stage, and gut compartment in the gut bacterial communities of two Cerambycidae species (Coleoptera). *J Microbiol.* **55**:21–30.
- Kirsch R, Gramzow L, Theissen G, Siegfried BD, Ffrench-Constant RH, Heckel DG, Pauchet Y. 2014. Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: key events in the evolution of herbivory in beetles. *Insect Biochem Mol Biol.* **52**:33–50.
- Kirsch R, Wielsch N, Vogel H, Svatos A, Heckel DG, Pauchet Y. 2012. Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. *BMC Genom.* **13**:587.
- Kuraku S, Zmasek CM, 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 Res.* **41**:W22–W28.
- Linsley EG. 1959. Ecology of Cerambycidae. *Annu Rev Entomol.* **4**:99–138.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**:459–473.
- McGrath CL, Gout J-F, Johri P, Doak TG, Lynch M. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res.* **24**:1665–1675.
- McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn SJ, Arsala D, 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 Biol.* **17**:227.
- McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE, Friedrich F, Letsch H, et al. 2019. The evolution and genomic basis of beetle diversity. *Proc Natl Acad Sci U S A.* **116**:24729–24737.
- Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* **30**:1188–1195.
- Monné ML, Monné MA, Wang Q. 2017. General morphology, classification, and biology of Cerambycidae. In: Wang Q, editors. *Cerambycidae of the world: biology and pest management*. Boca Raton (FL): CRC Press. p. 1–70.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* **32**:268–274.
- Pauchet Y, Kirsch R, Giraud S, Vogel H, Heckel DG. 2014. Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the Cerambycid beetle *Apriona japonica*. *Insect Biochem Mol Biol.* **49**:1–13.
- 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**:1517–1525.
- Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH. 2010. Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PLoS One* **5**:e15635.
- Santibanez L, Henriquez C, Corro-Tejeda R, Bernal S, Armijo B, Salazar O. 2021. Xylooligosaccharides from lignocellulosic biomass: a comprehensive review. *Carbohydr Polym.* **251**:117118.
- Schloss PD, Delalibera I, Handelsman J, Raffa KF. 2006. Bacteria associated with the guts of two wood-boring beetles: *Anoplophora glabripennis* and *Saperda vestita* (Cerambycidae). *Environ Entomol.* **35**:625–629.
- Scully ED, Hoover K, Carlson JE, Tien M, Geib SM. 2013. Midgut transcriptome profiling of *Anoplophora glabripennis*, a Lignocellulose degrading cerambycid beetle. *BMC Genom.* **14**:850.
- Shin NR, Shin S, Okamura Y, Kirsch R, Lombard V, Svacha P, Denux O, Augustin S, Henrissat B, McKenna DD, et al. 2021. Larvae of longhorned beetles (Coleoptera: Cerambycidae) have evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes. *Syst Entomol.* **46**:784–797.
- Sugimura M, Watanabe H, Lo N, Saito H. 2003. Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn Beetle, *Psacotheta hilaris*. *Eur J Biochem.* **270**:3455–3460.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**:W609–W612.
- Svacha P, Lawrence JF. 2014. Cerambycidae Latreille, 1802. In: Beutel RG and Leschen RAB, editors. *Handbook of zoology Arthropoda: Insecta, Coleoptera, beetles: morphology and systematics (Phytophaga)*. Berlin, Germany: Walter de Gruyter. p. 77–178.
- Vazquez MJ, Alonso JL, Dominguez H, Parajo JC. 2000. Xylooligosaccharides: manufacture and applications. *Trends Food Sci Technol.* **11**:387–393.

- Walsh B. 2003. Population-genetic models of the fates of duplicate genes. In: Long M, editors. *Origin and evolution of new gene functions*. Berlin: Springer. p. 279–294.
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**:W296–W303.
- Wei YD, Lee KS, Gui ZZ, Yoon HJ, Kim I, Zhang GZ, Guo X, Sohn HD, Jin BR. 2006. Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry Longicorn beetle, *Apriona Germari*. *Comp Biochem Physiol B: Biochem Mol Biol.* **145**:220–229.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* **13**:555–556.
- Zhu Y, Han L, Hefferon KL, Silvaggi NR, Wilson DB, McBride MJ, Spormann AM. 2016. Periplasmic *Cytophaga Hutchinsoii* endoglucanases are required for use of crystalline cellulose as the sole source of carbon and energy. *Appl Environ Microbiol.* **82**:4835–4845.