Larvae of longhorned beetles (Coleoptera; Cerambycidae) have evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes

Na Ra Shin, Seunggwan Shin, Yu Okamura, Roy Kirsch, Vincent Lombard, Petr Svacha, Olivier Denux, Sylvie Augustin, Bernard Henrissat, Duane Mckenna, et al.

Systematic Entomology

Systematic Entomology (2021), **46**, 784–797 DOI: 10.1111/syen.12488

Larvae of longhorned beetles (Coleoptera; Cerambycidae) have evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes

NA RA SHIN[1], SEUNGGWAN SHIN[2,3,4], YU OKAMURA[1], ROY KIRSCH[1], VINCENT LOMBARD[5,6], PETR SVACHA[7], OLIVIER DENUX[8], SYLVIE AUGUSTIN[8], BERNARD HENRISSAT[9], DUANE D. MCKENNA[2,3] and YANNICK PAUCHET[1]

[1]Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany, [2]Department of Biological Sciences, University of Memphis, Memphis, TN, U.S.A., [3]Center for Biodiversity Research, University of Memphis, Memphis, TN, U.S.A., [4]School of Biological Sciences, Seoul National University, Seoul, Republic of Korea, [5]CNRS, UMR 7257, Aix-Marseille Université, Marseille, France, [6]USC 1408 AFMB, INRAE, Marseille, France, [7]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic, [8]INRAE, URZF, Orléans, France and [9]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract. Longhorned beetles (Cerambycidae) are the most diverse group of predominantly wood-feeding (xylophagous) insects on Earth. Larvae of most species feed within tissues of plants made up of large amounts of plant cell wall (PCW), which is notoriously difficult to digest. To efficiently access nutrients from their food source, cerambycid larvae have to deconstruct PCW polysaccharides – such as cellulose, hemicelluloses and pectin – requiring them to possess a diversity of plant cell wall degrading enzymes (PCWDEs) in their digestive tract. Genomic data for Cerambycidae are mostly limited to notorious forest pests and are lacking for most of the taxonomic groups. Consequently, our understanding of the distribution and evolution of cerambycid PCWDEs is quite limited. We addressed the numbers, kinds and evolution of cerambycid PCWDEs by surveying larval midgut transcriptomes from 23 species representing six of the eight recognized subfamilies of Cerambycidae and each with very diverse host types (i.e., gymnosperms, angiosperms, xylem, phloem, fresh or dead plant tissues). Using these data, we identified 340 new putative PCWDEs belonging to ten carbohydrate active enzyme families, including two gene families (GH7 and GH53) not previously reported from insects. The remarkably wide range of PCWDEs expressed by Cerambycidae should allow them to break down most PCW polysaccharides. Moreover, the observed distribution of PCWDEs encoded in cerambycid genomes agreed more with phylogenetic relationship of the species studied than with the taxonomic origin or quality of the host plant tissues.

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

Introduction

Longhorned beetles (family Cerambycidae sensu stricto) form the largest family of predominantly wood-feeding (xylophagous) insects on Earth, with an estimated 5300 genera and 36 300 extant species (Monné *et al.*, 2017). They are distributed worldwide but are most species-rich in the tropics (Monné *et al.*, 2017). There are currently up to eight recognized subfamilies of Cerambycidae: Cerambycinae, Dorcasominae, Lamiinae, Lepturinae, Necydalinae, Parandrinae, Prioninae and Spondylidinae (Haddad *et al.*, 2018); Necydalinae and Parandrinae have been recently included in Lepturinae and Prioninae, respectively (Nie *et al.*, 2021). Lamiinae (∼21 000 species) and Cerambycinae (∼11 200 species) are the most species-rich subfamilies (Monné *et al.*, 2017). The family Cerambycidae is closely related to Chrysomelidae (leaf beetles; >35 000 species) and together with the much smaller families Disteniidae, Megalopodidae, Orsodacnidae, Oxypeltidae and Vesperidae make up the extraordinarily species-rich superfamily Chrysomeloidea (>60 000 species), itself sister to the superfamily Curculionoidea (weevils; >65 000 species) (Oberprieler *et al.*, 2007). Overall, Chrysomeloidea and Curculionoidea form the informal clade Phytophaga, which is, aside from Lepidoptera, the most diverse clade of predominantly plant-feeding (phytophagous) animals on Earth (Marvaldi *et al.*, 2009). The larvae of longhorned beetles are mostly specialized plant-feeders, often boring deep into woody plant tissues (Haack, 2017), and although larvae of some limited subgroups deviate biologically (terricolous species feeding externally on roots or even on fungus-infested plant debris, borers in annual herbs, in generative plant tissues, leaf miners, etc.), xylophagy *sensu lato* is very probably ancestral for the family because xylophagous species either strongly predominate or are the only known feeding type in all subfamilies. Many species play an important role in forest ecosystems by participating in the recycling of dead/dying wood, but some species, such as the Asian longhorned beetle, *Anoplophora glabripennis* (Motschulsky), are globally significant invasive pests capable of inflicting severe damage on living trees, including many economically important orchard, ornamental and forest tree species.

Most tissues of woody plants are predominantly composed of plant cell wall (PCW); either pectin-rich primary cell wall in living tissues, such as sapwood, cambium and inner bark, or lignocellulose-rich secondary cell wall in dead tissues, such as heartwood and outer bark. Woody plant tissues typically contain low amounts of nutrients essential to insect growth and development, including proteins, lipids and vitamins (Mattson Jr, 1980). Therefore, cerambycid larvae must efficiently break down the structural components – mostly polysaccharides – of the PCW to get access to as many nutrients as possible. To achieve efficient breakdown of these polysaccharides, such as cellulose, hemicelluloses (xylan, xyloglucan, β-glucans and mannans) and pectin, cerambycid larvae employ plant cell wall degrading enzymes (PCWDEs), some of which are self-produced, that is, secreted by the cells of their digestive tract (Hanks, 1999; Haack, 2017).

The presence of PCWDEs in the gut fluid of wood-boring cerambycid larvae was detected for the first time in the early 1900s (Linsley, 1959). Initially, cellulose-degrading enzymes were thought to be secreted by symbiotic yeasts, which in some groups reside in structures derived from the anterior midgut called the mycetomes (Buchner, 1928). Several studies from the 1930s challenged and eventually ruled out the fact that these enzymes originated from symbionts. These studies demonstrated that (i) species of Cerambycidae which did not harbour symbiotic yeasts were as efficient as species harbouring these symbionts in breaking down cellulose; and (ii) that symbiotic yeasts, when isolated and cultured, were not cellulolytic (Mansour & Mansour-Bek, 1934; Martin, 1983). It is now widely accepted that mostly symbiotic-independent cellulose and hemicellulose digestion occurs in cerambycid larvae and that the genome of these beetles encode their own PCWDEs (Scully *et al.*, 2013; Pauchet *et al.*, 2014; McKenna *et al.*, 2016, 2019).

A range of studies, mostly during 1970–1990, investigated the digestive abilities of cerambycid larvae (mostly Prioninae, Lepturinae, Cerambycinae and Lamiinae) with some of the responsible enzymes isolated. These studies showed that the gut of wood-boring larvae of Cerambycidae harboured both high pectolytic activity and activity against several hemicellulose components, including xylan, β-glucans and mannans. Although activity against amorphous cellulose was always detected, larvae were mostly unable to break down crystalline cellulose (Foglietti *et al.*, 1971; Chararas & Chipoulet, 1982, 1983; Chipoulet & Chararas, 1985a,b; Scrivener *et al.*, 1997). Interestingly, only the gut fluid of species whose larvae fed on decaying/rotten wood harboured enzymes able to digest crystalline cellulose. However in this case, the enzymes active on crystalline cellulose seem to originate from fungi present at the surface of decaying/rotten wood, and ingested by the cerambycid larvae when feeding, and not from the larvae themselves (Kukor & Martin, 1986; Kukor *et al.*, 1988).

Although the presence of PCWDEs in the gut fluid of wood-boring larvae of Cerambycidae has been known for a long time, the genes encoding them remained elusive until recently. Our knowledge about the molecular biology of Cerambycidae has been hampered mostly by the difficulty of rearing them in the lab and by their long development time. The first gene encoding a cerambycid-derived PCWDE – an endo-β-1,4-glucanase from subfamily 2 of the glycoside hydrolase family 5 (GH5_2) – was cloned from the yellow spotted longhorned beetle (*Psacothea hilaris* (Pascoe)) (Sugimura *et al.*, 2003). Most of what is currently known about the function of genes encoding PCWDEs in longhorned beetles is restricted to pest species in the subfamily Lamiinae (Lee *et al.*, 2004, 2005; Pauchet *et al.*, 2014; Liu *et al.*, 2015; McKenna *et al.*, 2016). Additionally, whole body transcriptomes sequenced from adult specimens of Cerambycinae, Lepturinae and Parandrinae recently became available (McKenna *et al.*, 2019). Based on these data, genomes of Cerambycidae appear to regularly encode putative cellulases from families GH9, GH45, GH48 and GH5_2 and polygalacturonases (PG, pectinases) from the family GH28.

To provide new insights into the distribution of PCWDEs and their evolutionary history within Cerambycidae as a whole, we

sequenced larval midgut transcriptomes of 23 species mostly from the western Palearctic region, belonging to six of the eight cerambycid subfamilies. Using these data, we first reconstructed the phylogeny of the corresponding species. Then, we exhaustively searched these transcriptomes and annotated 340 transcripts encoding new putative PCWDEs. We found that Cerambycidae express a remarkably diverse array of putative PCWDEs, indicating that their larvae likely possess the ability to digest most structural polysaccharides composing PCWs.

Materials and methods

Insect collection

Specimens corresponding to 21 species were collected from the field in France, Czech Republic and Slovakia. Our goal was to obtain representative species spanning as many higher-level groups of Cerambycidae, and as wide a spectrum of feeding types as possible. However, we were restricted by the fact that we needed actively feeding, freshly caught larvae to ensure the extraction of high-quality RNA for sequencing. The larvae were sent alive to the Max Planck Institute for Chemical Ecology (Jena, Germany) with original food source material. *Hylotrupes bajulus* (Audinet-Serville) originated from a lab culture kept at the 'Materialprüfanstalt Brandenburg GmbH' (Eberswalde, Germany) and then larvae were sent to the Institute in cut pieces of pine softwood. *Cacosceles newmannii* (Thomson) originated from a lab culture kept at the Centre for Invasion Biology, Stellenbosch University (South Africa) (Table 1). For this species, dissected guts were sent to the lab preserved in RNAlater. Larvae collected from the field were identified primarily using morphological characters, combined with data on host plant association. In addition, identity of specimens was confirmed by comparing the cytochrome oxidase I (COI) barcode extracted from the assembled transcriptomes (see later) to known COI barcodes deposited in GenBank (http://www.ncbi.nlm.nih.gov) and Barcode of Life Data System version 4 (v4BOLD; https://v4.boldsystems.org/).

RNA extraction and sequencing

For specimens sent alive to the lab, larvae were first surface-sterilized by briefly dipping them in 70% ethanol before extensive rinsing in double-distilled water. Larvae were then cooled down on ice for 10 min before being immobilized on a dissecting surface using needles. Larvae were opened ventrally to expose the digestive tract and the midgut was removed. Mycetomes, if present, were removed as carefully as possible. Dissected guts were then extensively washed in sterile water to remove their contents, before being snap frozen in liquid nitrogen and stored at −80°C until further use. Only actively feeding larvae, checked by the presence of a food bolus inside the digestive tract, were selected for dissection. Total RNA was extracted from the dissected midgut of a single specimen per species – in order to reduce heterozygosity in the resulting total RNA samples – using TRIzol (Thermo Fisher Scientific, Darmstadt, Germany) following the manufacturer's procedure. Potential genomic DNA contamination was removed by DNase treatment (TURBO DNase, Thermo Fisher Scientific) for 30 min at 37°C. Total RNA samples were further purified using the RNeasy MinElute Cleanup Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol and eluted in 20 μL of RNA storage solution (Thermo Fisher Scientific). Integrity and quality of RNA samples were determined using the RNA 6000 Nano LabChip kit on an Agilent 2100 Bioanalyzer (both from Agilent Technologies, Santa Clara, CA, U.S.A.) following the manufacturer's guidelines. RNA-Seq was outsourced to the Max Planck-Genome centre (Cologne, Germany) and to GATC Biotech (https://www.gatc-biotech.com, Konstanz, Germany). Poly(A)+-RNA was enriched before being fragment to an average of 300–350 nucleotides. Paired-end sequencing libraries (2×150 bp) were then generated using the TruSeq Stranded mRNA Library Prep kit (Illumina, San Diego, CA, U.S.A.) and single-indexed adapter tags (double-indexed adapter tags for *C. newmannii*). The resulting libraries were sequenced on an Illumina HiSeq 2500 platform.

RNA-Seq data preprocessing and assembly

Adapter sequence and low-quality bases were removed from the resulting sequencing reads using Trimmomatic v0.36 (Bolger *et al.*, 2014) with the following parameters: slidewindow: 4:5; leading: 5; trailing: 5; minlen: 25. Transcriptome assembly was performed using Trinity v2.5.1 (Grabherr *et al.*, 2011) with the following parameters: minimum contig length 250 and minimum k-mer coverage 1. Since (i) some of the sequencing libraries were multiplexed on the same sequencing lanes, and that (ii) single-indexed tags were used, extensive cross-contaminations between several transcriptome datasets occurred after read de-multiplexing. In order to eliminate these cross-contaminated reads, we used a protocol described by Peters *et al.* (2017). After transcriptome assembly, we performed reciprocal BLAST searches, using BLASTn, between assemblies of transcriptome datasets, which were sequenced on the same lane or even during the same run. Transcripts that shared nucleotide sequence identity of at least 98% over a length of at least 200 bp between two or more assemblies were identified. If the relative coverage of two transcripts originating from two different assemblies differed >twofold, the transcript with the lower relative coverage was assumed to be a contaminant and was removed from the corresponding assembly. The raw data and assembled transcriptomes (BioProject ID: PRJNA609440) are deposited in the corresponding National Center for Biotechnology Information (NCBI) databases (ncbi_sra and ncbi_tsa) under the accession numbers provided in Supporting data 1, Table S5.

Species phylogeny

We generated a phylogenetic tree based on the transcriptome data generated from our 23 cerambycid taxa as well

Table 1. Summary of the cerambycid species analyzed in this study.

Taxon name	Subfamily	Tribe	Host plant	Location
Aegosoma scabricorne	Prioninae	Aegosomatini	Decaying poplar logs (*Populus nigra*)	commune de Baule (Loiret, 45), France
Prionus coriarius	Prioninae	Prionini	Collar of hornbeam tree (*Carpinus* sp.)	Forêt domaniale d'Orléans (Loiret, 45), France
Ergates faber	Prioninae	Ergatini	Decaying pine trunk	Massif des Maures (83), France
Cacosceles newmannii	Prioninae	Cacoscelini	Living sugarcane stalks (*Saccharum* sp.)	lab culture, Stellenbosch University, South Africa
Phymatodes testaceus	Cerambycinae	Callidiini	Oak (*Quercus* sp.) under the bark	Forêt domaniale de Bellême (Orne, 61), France
Plagionotus detritus	Cerambycinae	Clytini	Oak (*Quercus* sp.) under the bark	Forêt domaniale d'Orléans (Loiret, 45), France
Hylotrupes bajulus	Cerambycinae	Hylotrupini	Softwood pine timber	lab culture Eberswalde, Germany
Molorchus minor	Cerambycinae	Molorchini	Under bark of broken dead Spruce (*Picea* sp.)	Bohemia, České Budějovice, Czech Republic
Aromia moschata	Cerambycinae	Callichromatini	In living thin stems of *Salix caprea*	Bohemia, Krusne Hory Mts., Kovarska (850 m), Czech Republic
Rhagium bifasciatum	Lepturinae	Rhagiini	Decayed beech wood	Forêt domaniale de Bellême (Orne, 61), France
Leptura aurulenta	Lepturinae	Lepturini	Decaying oak wood	Forêt domaniale de Bellême (Orne, 61), France
Carilia virginea	Lepturinae	Rhagiini	Under loose bark of dead fallen thin spruce (*Picea* sp.)	Bohemia, Krusne Hory Mts., Kovarska (850 m), Czech Republic
Oxymirus cursor	Lepturinae	Oxymirini	Wood of rotten *Picea* log on the ground	Bohemia, Cesky Krumlov env., Klet Mt., Czech Republic
Rhamnusium bicolor	Lepturinae	Rhamnusiini	Moist basal part of thick dead branch on large living *Populus* tree	Sturovo env., Pavlova, Slovakia
Necydalis major	Necydalinae	Necydalini	Rotting wood of dead standing *Alnus glutinosa*	Bohemia, České Budějovice env. (Kaliste near Lipi), Czech Republic
Acanthocinus aedilis	Lamiinae	Acanthocinini	Under bark of large fallen *Pinus* stem	Bohemia, Ceske Budejovice env. (Kaliste near Lipi), Czech Republic
Saperda scalaris	Lamiinae	Saperdini	Under bark of standing dead *Alnus glutinosa*	Bohemia, České Budějovice env. (Nove Homole), Czech Republic
Mesosa nebulosa	Lamiinae	Mesosini	In wood of dry dead standing small oak (*Quercus* sp.)	Sturovo env., Chlaba, Slovakia
Exocentrus adspersus	Lamiinae	Pogonocherini	Dry oak branches	Sturovo env., Chlaba, Slovakia
Agapanthia villosoviridescens	Lamiinae	Agapanthiini	Stems of dead Apiaceae (probably *Conium maculatum*)	Bohemia, České Budějovice, Czech Republic
Anisarthron barbipes	Spondylidinae	Anisarthrini	In dead moist parts of living small *Ulmus* trees	Sturovo env., Hegyfarok, Slovakia
Saphanus piceus	Spondylidinae	Saphanini	Rotting wood at or under ground level in dead standing stems of *Corylus avellana*	Bohemia, Cesky Krumlov env., Czech Republic
Spondylis buprestoides	Spondylidinae	Spondylidini	Root of a large *Pinus* stump	Bohemia, České Budějovice env. (Kaliste near Lipi), Czech Republic

Taxon names were attributed according to morphological characters and DNA barcoding.

as eight Coleoptera from the 1KITE initiative (2 Chrysomelidae, 4 Curculionidae, 1 Byturidae and 1 Passalidae; the latter two were used as outgroup) (Misof *et al.*, 2014; McKenna *et al.*, 2019). We coded a custom pipeline to generate a phylogenetic tree based on a nucleotide (codon-based) alignment of detected BUSCO genes (Waterhouse *et al.*, 2018) from transcriptome data. First, the longest open reading frame (ORF) for each transcript was predicted using transDecoder (https://github.com/TransDecoder/TransDecoder/wiki). Then, a BUSCO analysis was performed for each dataset using the insecta_odb10 database (1658 genes), and BUSCO genes which were shared across all taxa were identified. The corresponding amino acid (AA) and nucleotide (NT) sequences were then extracted. Amino acid sequences were aligned in MAFFT v7.402 with – auto and default options. The resulting alignment was then back-translated into the corresponding NT alignment using pal2nal.pl (Suyama *et al.*, 2006), and gaps were trimmed from the alignment using trimAl 1.2rev59, accepting 5% gaps for each position (Capella-Gutierrez *et al.*, 2009). After concatenating aligned NT sequences for each taxon, we performed a maximum likelihood (ML) phylogenetic analysis using IQ-TREE 2.0.3 (Nguyen *et al.*, 2015). Statistical measures of nodal support were estimated using ultrafast bootstrap analysis (Hoang *et al.*, 2018) implemented in the IQ-TREE software

with 1000 replicates. For more details, please refer to Supporting data 2.

Screening for carbohydrate active enzymes and PCWDEs

We screened assembled larval midgut transcriptomes using the CAZy pipeline (Lombard *et al.*, 2014) to identify the set of carbohydrate active enzymes (CAZymes) in each transcriptome. This analysis was performed using the FASTY tool (Pearson *et al.*, 1997) against a sequence library derived from the CAZy database (http://www.cazy.org). The FASTY results were analysed using an internal script. Only the sequences that had an e-value better than 10^{-6} and an identity percentage $\geq 30\%$ were kept. In parallel, contigs encoding putative PCWDEs were identified through tBLASTn searches of the transcriptomes with an E-value cut-off 10^{-5} using amino acid sequences of PCWDEs previously identified in species of Phytophaga beetles. In addition, CAZyme families were screened according to their predicted function to assess whether new PCWDE families that were not previously described in insects were present. The corresponding contigs were retrieved and manually curated. Contigs found to encode only a partial ORF were used to design specific primer pairs to perform 5′- and 3′-Rapid Amplification of cDNA ends (RACE) PCRs (Supporting data 1; Table S6). For these, we used the SMARTer RACE cDNA Amplification kit (Takara Bio Inc, Mountain View, CA, U.S.A.) following the manufacturer's instructions. The resulting PCR products were cloned into the pCR4 TOPO/TA vector (Thermo Fisher Scientific), transformed in *Escherichia coli* TOP10 cells and Sanger sequenced. The resulting nucleotide sequences corresponding to the ORFs can be found in Supporting data 3.

Amino acid sequence alignments and phylogenetic analyses

Amino acid sequences curated from our transcriptomes and their counterparts from other animal and microbes extracted from NCBI were aligned as described earlier. We eliminated redundancy at 90% identity level from each sequence dataset using the CD-HIT Suite server (Huang *et al.*, 2010). Sequences were aligned with MAFFT (v.1.3.7) using its default settings, in Geneious 11.1.4. ML phylogenetic analyses were inferred using IQ-TREE (Nguyen *et al.*, 2015). The best-fit substitution model for each dataset was automatically determined using ModelFinder implemented in IQ-TREE. Branch supports were assessed using an ultrafast bootstrap approximation (1000 replicates) (Hoang *et al.*, 2018). For more details, please refer to Supporting data 4.

Draft genome assembly

We downloaded sequencing datasets corresponding to the draft genomes of two species of Cerambycinae (*Phymatodes lengi*: SRX3924042, SRX3924041; and *Xylotrechus colonus*: SRX3924050, SRX3924049) from the sequence read archive (SRA) at the NCBI (https://www.ncbi.nlm.nih.gov). Assembly was performed with CLC Genomics Workbench 10.0 (Qiagen) using default parameters. For more details, please refer to Supporting data 5. The resulting genome assemblies were mined for the presence of PCWDE-encoding genes using tBLASTn searches. A single representative sequence for each PCWDE gene family identified from the transcriptome of *Phymatodes testaceus* was used as BLAST query. Contigs harbouring PCWDE-encoding genes according to these BLAST searches were extracted for further analyses. Spliced sequence alignments were performed using the online program Splign (Kapustin *et al.*, 2008) to assess the intron/exon structure of the genes identified. Relaxed parameters were used to allow for cross-species analyses.

Results

Transcriptome sequencing and species phylogeny

We collected larval specimens of 23 species of Cerambycidae, mostly from Europe, representing 22 tribes and all recognized subfamilies, apart from Dorcasominae and Parandrinae (Table 1), and assembled their larval gut transcriptomes (Supporting data 1; Table S1). Species identifications based on morphological characters were confirmed by extracting the cytochrome oxidase I (COI) barcodes from the transcriptomes and comparing them to public databases by BLAST (Supporting data 1; Table S2). No COI barcode was available for *C. newmannii* for comparison in either GenBank or v4BOLD. The resulting larval midgut transcriptomes harboured various levels of completeness according to a Benchmarking Universal Single-Copy Orthologs (BUSCO; (Waterhouse *et al.*, 2018)) analysis (Supporting data 1; Table S3). The BUSCO analysis identified a low of 49.1% complete BUSCO genes in the transcriptome from *Acanthocinus aedilis* (Lamiinae) and a high of 84.9% complete BUSCO genes in the transcriptome from *C. newmannii* (Prioninae).

We used these transcriptomes to build a species phylogeny. We extracted BUSCO genes that were common between all transcriptomes, including taxa used as outgroups. A total of 124 BUSCO genes were recovered, their nucleotide sequences were concatenated for each transcriptome, and the resulting codon-based superalignment (81 990 sites) was used to perform an ML phylogenetic analysis (Supporting data 2 and Fig. 1). Subfamilies within Cerambycidae were all monophyletic and had maximum bootstrap support, except for Lepturinae, which included Necydalinae; the following clades were recovered: (1) Cerambycinae; (2) Prioninae; (3) Lepturinae incl. Necydalinae; (4) Spondylidinae, and (5) Lamiinae. The first split was between clade $(1 + 2)$ and $(3 + (4 + 5))$ and was maximally supported. The split between (3) and $(4 + 5)$ was strongly supported with 97% ML bootstrap support. The split between (1) and (2) was maximally supported, and that between (4) and (5) had 97% ML bootstrap support (Fig. 1). Overall, our results were consistent with the results of Haddad *et al.* (2018) for the subfamilies of Cerambycidae, except for clade $(3 + (4 + 5))$. In Haddad

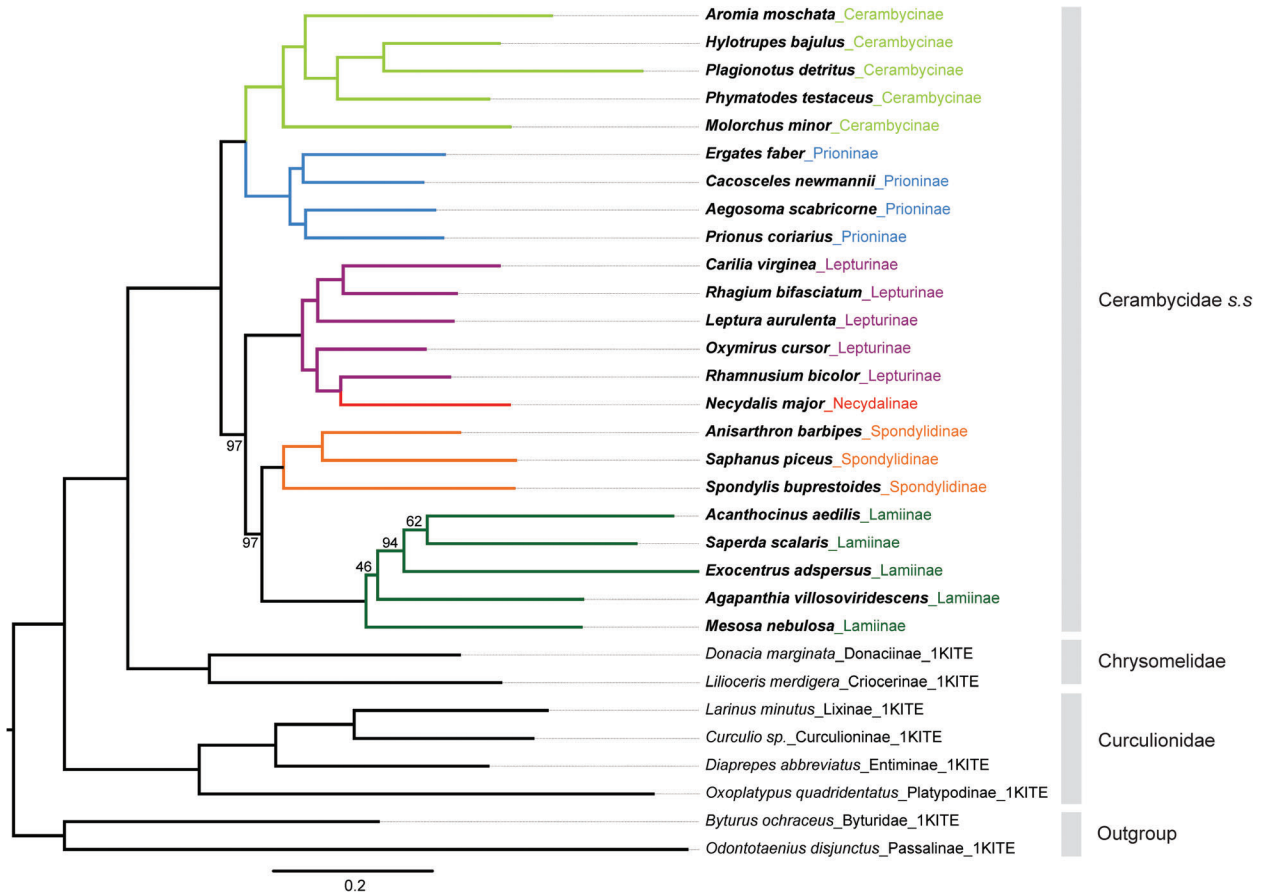


Fig. 1. Phylogenetic relationships between the species of Cerambycidae analysed in the present study. A maximum likelihood analysis was performed with 1000 ultrafast-bootstrap replicates using the program IQ-TREE. The analysis was based on a codon-based nucleotide sequence alignment of 124 BUSCO genes found to be common to all datasets used here. The sequence alignment contained 81 990 sites. The best model of nucleotide evolution was determined in IQ-TREE and was the generalized time reversible (GTR) model, with empirical base frequencies count (+F) and FreeRate model (+R6). Support values for nodes which were not maximally supported are indicated. Transcriptome from several beetle species from the 1KITE project were used as an outgroup (black branches): *Lilioceris merdigera* (Chrysomelidae; Criocerinae; GDPT00000000.1); *Donacia marginata* (Chrysomelidae; Donaciinae; GDMV00000000.1); *Curculio* sp. (Curculionidae; Curculioninae; GDNP00000000.1); *Diaprepes abbreviatus* (Curculionidae; Entiminae; GDNL00000000.1); *Larinus minutus* (Curculionidae; Lixinae; GDLA00000000.1); *Oxoplatypus quadridentatus* (Curculionidae; Platypodinae; GDNR00000000.1); *Byturus ochraceus* (Byturidae; GDLG00000000.1); *Odontotaenius disjunctus* (Passalidae; Passalinae; GDMS00000000.1). The branch containing the latter two species was used to root the tree. Branches are coloured by subfamily: Cerambycinae (light green), Prioninae (blue), Lepturinae (purple), Necydalinae (red), Spondylidinae (orange) and Lamiinae (green). More details about the data used to generate this species tree are found in Supporting data 2.

et al. (2018), clade (3) (Lepturinae, including Necydalinae) was associated with a clade formed by Prioninae and Cerambycinae (1 + 2). Also, Necydalinae was an ingroup of Lepturinae in our study, in contrast to Haddad *et al.* (2018), who recovered Necydalinae as a sister group of Lepturinae. In fact, the relationships between subfamilies of Cerambycidae we retrieved here were compatible with the incompletely resolved scheme proposed by Svacha & Lawrence (2014).

CAZymes and PCWDEs

We screened our cerambycid transcriptomes for the presence of CAZymes without discrimination according to their putative function (Supporting data 1; Table S4). Out of all the CAZyme families that were identified from this initial screen, those encoding putative PCWDEs were extracted. We first selected CAZyme families of PCWDEs already known from species of Phytophaga beetles, such as the cellulolytic GH9, GH5_2, GH45 and GH48; the pectinolytic GH28; and the hemicellulolytic GH5_8, GH10 and GH11. Other families of PCWDEs uncommon to Phytophaga beetles were then analysed, such as the putative cellulolytic GH7, and other subfamilies of GH5, together with putative debranching enzymes acting on hemicellulose or pectin, such as GH43_26 and GH53. Some of these CAZyme families were subsequently excluded because their sequences shared more than 95% amino acid identity with known microbial sequences after BLAST searches, and most likely originated

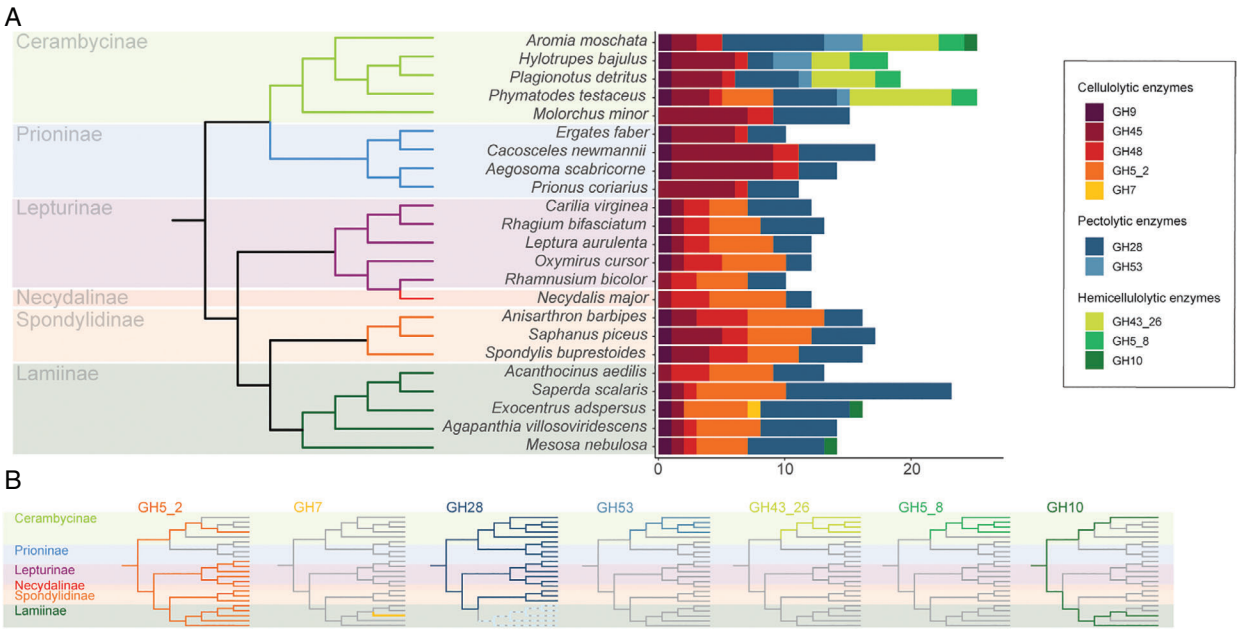


Fig. 2. Distribution of families of plant cell wall degrading enzymes in Cerambycidae. (A) The number of manually curated transcripts for each family of PCWDEs is put in the context of the species phylogeny. Each PCWDE family is represented by one colour (see insert for details). (B) For PCWDE families that were found not to be common to all species analysed, the distribution pattern is indicated. In the case of GH28, the distribution of the 'Lamiinae-type' ones is represented by dashed branches. Nucleotide sequences corresponding to the open reading frames of these PCWDEs are found in Supporting data 3.

Table 2. Summary of the PCWDE-encoding genes present in the shallow draft genomes of *Phymatodes lengi* and *Xylotrechus colonus*.

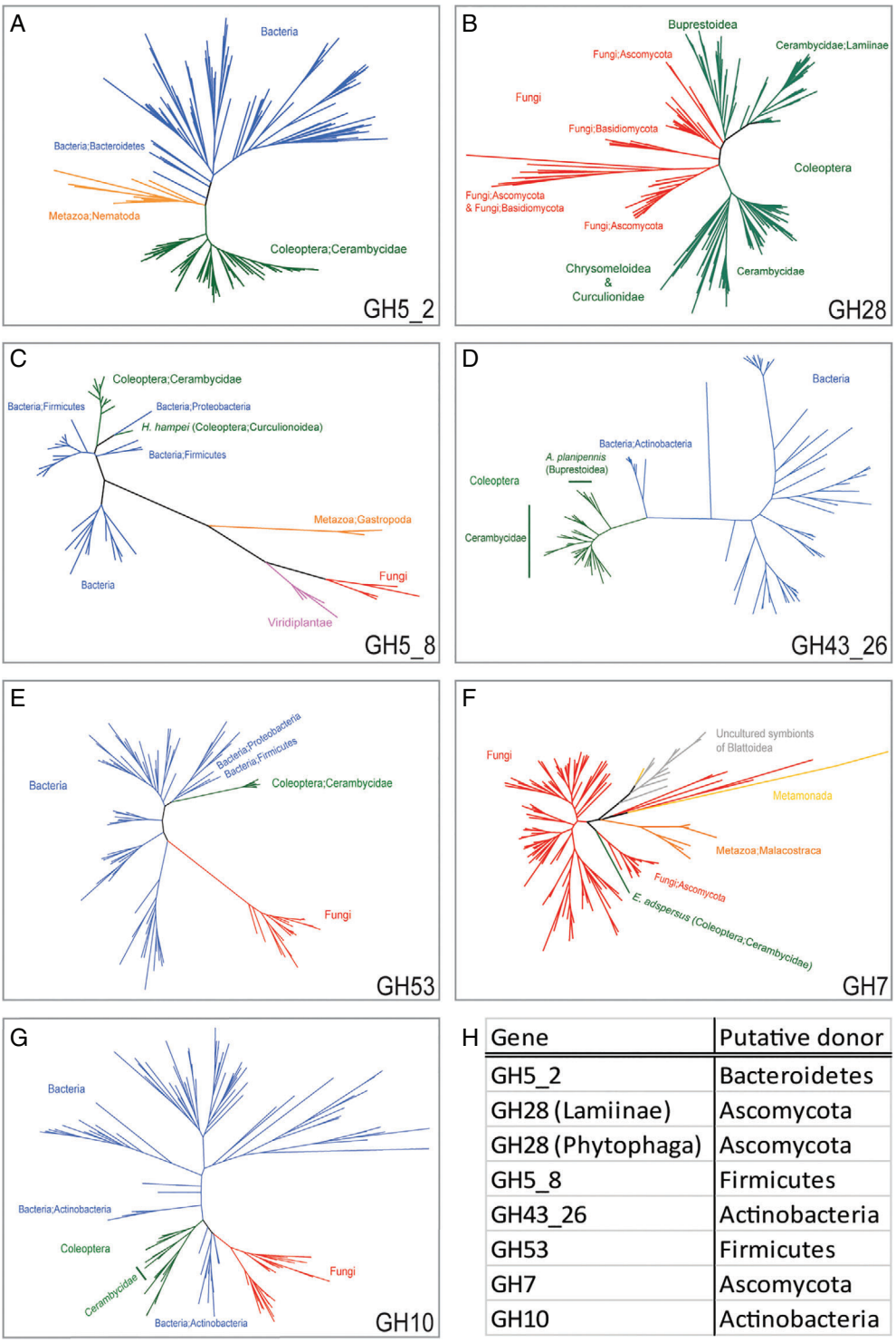
	GH9	GH45	GH48	GH5_2	GH7	GH28	GH53	GH5_8	GH10	GH43_26
P. lengi	1 (8)	4 (2/3)	1 (0)	3 (0)	–	11 (4)	1 (3)	2 (2)	–	15 (2)
X. colonus	1 (8)	4 (2/3)	1 (?)	1 (0)	–	12 (4)	1 (3)	3 (2)	1 (3)	11 (2)

For each family, the number of predicted genes and the corresponding number of predicted exons (brackets) is indicated.

from microbial contaminants, some of which could be part of the insect's microbiota. Altogether, we identified 340 new putative PCWDEs (Supporting data 3) expressed by species of Cerambycidae – all GHs – representing ten CAZyme families (Fig. 2A). The number of unique genes per species, based on our transcriptomes, ranged from a low of ten in *Ergates faber* (Prioninae) and *Rhamnusium bicolor* (Lepturinae), up to 26 in *P. testaceus* (Cerambycinae) (Fig. 2A). We plotted the distribution of these CAZyme families according to the phylogeny of the species investigated here (Fig. 2B) and realized that their distribution, apart from GH10 and GH7 (the latter being uniquely found in *Exocentrus* of Lamiinae), was phylogenetically conserved, with clades of longhorned beetle species sharing similar arrays of PCWDEs.

The cellulolytic GH9, GH45 and GH48 were distributed in all species investigated apart from GH9, which was missing from five species scattered rather widely across the phylogeny (Fig. 2A). Although we could not find a contig encoding a GH9 in our larval midgut transcriptome for *R. bicolor*, one could be found in the 1KITE adult beetle transcriptome for this species (GenBank: GDPK01005231.1). Thus, it is likely that all species of Cerambycidae investigated here harbour a GH9 gene in their genome, even though we did not find the corresponding transcripts in all of the gut transcriptomes studied. Consequently, we propose that these three gene families were present in the most recent common ancestor (MRCA) of the family Cerambycidae.

The cellulolytic GH5 subfamily 2 (GH5_2) (Aspeborg *et al.*, 2012) was widely distributed in the species investigated here, except for Prioninae and most Cerambycinae (Fig. 2A). We found members of this gene family in a single Cerambycinae transcriptome (*P. testaceus*), and in the genome data of *P. lengi* and *Xylotrechus colonus* (both Cerambycinae) we retrieved from Genbank (Table 2). A gene family-specific phylogenetic analysis indicated that cerambycid-derived GH5_2 sequences were monophyletic and cluster next to a clade of Nematoda-derived and Bacteroidetes-derived counterparts (Fig. 3A). According to their distribution within Cerambycidae and beyond, we conclude that GH5_2 genes may have been horizontally acquired by the Cerambycidae MRCA from a Bacteroidetes donor (as in McKenna *et al.*, 2019) and evolved to become a multigene family in a large number of longhorned beetle species; however, some groups eventually lost these genes



H

Gene	Putative donor
GH5_2	Bacteroidetes
GH28 (Lamiinae)	Ascomycota
GH28 (Phytophaga)	Ascomycota
GH5_8	Firmicutes
GH43_26	Actinobacteria
GH53	Firmicutes
GH7	Ascomycota
GH10	Actinobacteria

Fig. 3. Horizontal acquisition of PCWDEs in species of Cerambycidae. The phylogenetic relationships between cerambycid-derived PCWDE families and their microbial counterparts was analysed using maximum likelihood. Analyses were performed in IQ-TREE with 1000 ultrafast-bootstrap replicates. (A) GH5_2; (B) GH28; (C) GH5_8; (D) GH43_26; (E) GH53; (F) GH7; (G) GH10; (H) summary table of the putative microbial donors. More details regarding data that were used to generate these trees are found in Supporting data 4.

during the course of evolution, especially within Cerambycinae and Prioninae (Fig. 2B).

We identified two types of pectinolytic GH28 sequences in the 23 cerambycid transcriptomes we generated (Fig. 2B). First, the so-called 'Phytophaga-ancestral' type of GH28 gene was acquired by the MRCA of Phytophaga beetles from a Pezizomycotina fungus donor through a horizontal gene transfer (HGT) event (Fig. 3B; Kirsch *et al.*, 2014). This type of GH28 protein is characterized by several synapomorphic amino acids (Kirsch *et al.*, 2014), such as the presence of ten cysteine residues forming five disulphide bridges as well as a highly conserved histidine residue, which is substituted to an arginine residue in most fungal sequences (Fig. S1). This 'Phytophaga-ancestral' type of GH28 was distributed in all subfamilies of Cerambycidae we investigated apart from Lamiinae. In the latter subfamily, all species harboured what we call here the 'Lamiinae-type' of GH28 (Fig. S2), which was most likely originally acquired by the MRCA of Lamiinae from a Saccharomycotina fungus donor through HGT, according to our phylogenetic analysis (Fig. 3B). This type of GH28 protein is characterized by the presence of only three disulphide bridges and the highly conserved arginine residue, instead of histidine, which is typically found in the 'Phytophaga-ancestral' type of GH28 (Kirsch *et al.*, 2014; Pauchet *et al.*, 2014; McKenna *et al.*, 2016). According to their distribution, we argue that the genome of the Cerambycidae MRCA encoded the 'Phytophaga-ancestral' type of GH28. After the split which gave rise to the Lamiinae, the MRCA of the latter subfamily lost the 'Phytophaga-ancestral' type of GH28 and acquired the 'Lamiinae-type' one by HGT, most likely from a Saccharomycotina fungus donor (Fig. 2B).

Transcripts encoding GH5 subfamily 8 (GH5_8), GH43 subfamily 26 (GH43_26) and GH53 were restricted to species of Cerambycinae, except *Molorchus minor* (L.) (Fig. 2A). GH5_8 proteins are typical endo-acting mannanases (EC 3.2.1.78) implicated in the degradation of β-mannan (Aspeborg *et al.*, 2012). According to a phylogenetic analysis (Fig. 3C), cerambycid-derived GH5_8 proteins were recovered as monophyletic and cluster next to Firmicutes counterparts.

GH43_26 proteins, which have been functionally characterized so far, are debranching enzymes and possess α-L-arabinofuranosidase activity (EC 3.2.1.55). They play a role in the degradation of arabinoxylan by removing the arabinose decorations attached to the backbone of this polysaccharide (Mewis *et al.*, 2016). A phylogenetic reconstruction of the evolution of this family of proteins recovered cerambycid-derived GH43_26 as monophyletic, sister to those derived from the jewel beetle *Agrilus planipennis* (Fairmaire) (Buprestidae), and close to a clade of Actinobacteria counterparts (Fig. 3D), consistent with the results of McKenna *et al.* (2019).

To date, the sole function attributed to GH53 proteins is endo-β-1,4-galactanase (EC 3.2.1.89) implicated in the breakdown of galactans and arabinogalactans, two polysaccharides of the pectin network, forming the side chains of rhamnogalacturonan I (http://www.cazy.org/GH53.html). In our phylogenetic analysis, cerambycid-derived GH53 sequences were monophyletic and sister to a clade of Firmicutes-derived GH53 sequences (Fig. 3E).

As mentioned earlier, these three gene families appear to be restricted to species of Cerambycinae and may not have been present in the MRCA of Cerambycidae (Fig. 2B), but may have been acquired by the MRCA of the subfamily Cerambycinae from bacterial donors through one or several HGT events.

GH7 transcripts corresponding to a single gene were detected in only one species, the lamiine *Exocentrus adspersus* (Mulsant) (Fig. 2A). GH7 proteins are mostly considered cellulolytic, acting as endo-β-1,4-glucanase (EC 3.2.1.4) or as reducing end-acting cellobiohydrolase (EC 3.2.1.176). This is the first description of a member of this gene family in Phytophaga beetles and in insects in general. Results of our phylogenetic analysis indicated that the *E. adspersus* GH7 sequence clusters next to Pezizomycotina counterparts (Fig. 3F). We speculate that this gene has been acquired recently in the evolution of Cerambycidae (Fig. 2B), but may not be restricted to *E. adspersus*. Similarly, transcripts encoding GH10 putative xylanases (EC 3.2.1.8) were detected in several species scattered across the tree (Fig. 2A). These GH10 sequences were recovered as monophyletic by phylogenetic analysis next to a cluster of Actinobacteria-derived sequences (Fig. 3G). These genes may have been present in the MRCA of the Cerambycidae followed by multiple losses during the diversification of this beetle family (Fig. 2B).

Genomic organization of Cerambycidae PCWDEs

Although the presence of gene families encoding GH9, GH45, GH48 and 'Lamiinae-type' GH28 proteins has been already confirmed in the genome of the Asian longhorned beetle, *A. glabripennis* (McKenna *et al.*, 2016), this is not yet the case for other families of PCWDEs discovered during our transcriptome analysis, such as GH5_8, GH43_26 and GH53. Considering their resemblance to microbe-derived sequences (Fig. 3), we asked whether genes encoding these enzymes were indeed endogenous to cerambycid genomes. To address this point, we took advantage of shallow genome sequencing projects deposited in GenBank from *P. lengi* (Joutel) and *Xylotrechus colonus* (Fabricius), two species of Cerambycinae. We generated genome assemblies for both species using the raw reads from NCBI (Supporting data 5) and screened them for the presence of PCWDEs. We confirmed the presence of all of the gene families mentioned earlier (Table 2). GH53 genes seemed to be present as single copy in both shallow genomes, whereas GH5_8 and GH43_26 formed multigene families (Table 2). Genes encoding GH5_8 and GH43_26 are composed of two exons, whereas the GH53 genes contained three exons (Table 2 and Fig. 4). GH43_26 seemed to have mostly evolved through tandem gene duplication events (Fig. 4).

Discussion

Phytophagous longhorned beetles possess a battery of enzymes conveying the ability to break down most components of the PCW (Martin, 1983, 1991; Haack, 2017). To date, the genes

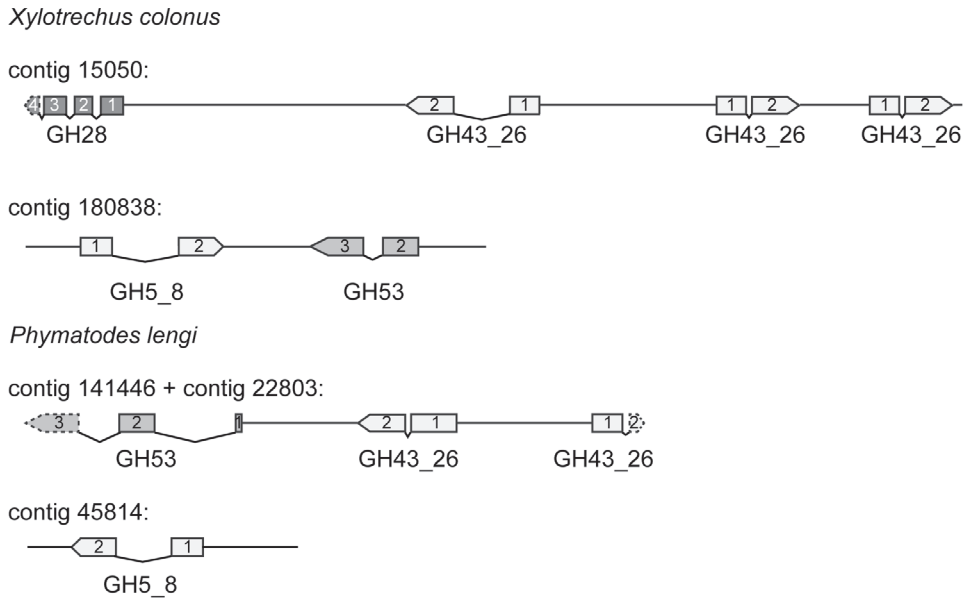


Fig. 4. Structure and organization of genes encoding PCWDEs in *Xylotrechus colonus* and *Phymatodes lengi*. Short read datasets corresponding to shallow genome sequencing projects of the two Cerambycinae species *X. colonus* and *P. lengi* were retrieved from GenBank and assembled. Selected contigs containing genes encoding representatives of several PCWDE families are depicted for *X. colonus* and *P. lengi*. Exons are indicated by boxes in various shades of grey for each gene. Dashed boxes indicate that a portion of the sequence is missing. Individual exons are numbered for each gene.

encoding some of these enzymes – especially cellulases – have been cloned and sequenced, but the species targeted were forest pests belonging to the subfamily Lamiinae (Sugimura *et al.*, 2003; Lee *et al.*, 2004, 2005; Wei *et al.*, 2006; Mei *et al.*, 2016). Transcriptome and genome surveys have revealed that these genes, in species of Lamiinae, were part of medium-size gene families rather than being single copy, and confirmed that these genes were endogenous to cerambycid beetle genomes (Scully *et al.*, 2013; Pauchet *et al.*, 2014; Liu *et al.*, 2015; McKenna *et al.*, 2016). It has since become clear that cellulolytic GH5_2, GH9, GH45 and GH48, as well as pectinolytic GH28, form the complement of PCWDEs encoded in the genomes of Cerambycidae. Our results show that this picture is more complex than previously thought. For example, although widely represented in our 23 transcriptomes, GH5_2 is not a hallmark of Cerambycidae as a whole, as shown by their absence in most of the Cerambycinae and in all the Prioninae transcriptomes investigated here, and consistent with the results of McKenna *et al.* (2019). We also confirm recent findings (McKenna *et al.*, 2019) that mannanolytic GH5_8, xylanolytic GH10 and GH43_26 belong to the arsenal of PCWDEs of Cerambycidae. Moreover, we describe for the first time in insects novel families of PCWDEs, including cellulolytic GH7 and pectinolytic GH53 genes. Despite their strong resemblance to bacterial-derived sequences, we confirmed that genes encoding GH5_8, GH43_26, GH53 and GH10 were endogenous to cerambycid beetle genomes, after inspecting shallow genome datasets derived from two species of Cerambycinae. Although we found a single GH7 in the lamiine species *E. adspersus*, this gene family may be distributed more widely in Cerambycidae, as suggested by a BLAST search in which the top hit corresponded to a GH7 sequence derived from another lamiine, *Mesosa myops* (AMD09875.1; 88% amino acid identity).

The fact that PCWDEs are part of multigene families opens up the possibility of functional diversification through change in substrate specificity or sub/neofunctionalization. If this is the case, this would mean that members of a single CAZyme family could break down a set of PCW polysaccharides rather that a single one. This would reduce the apparent complexity of the PCWDE arsenal – in terms of number of CAZyme families – for a given species, without reducing its ability to digest PCW polysaccharides. Such a phenomenon has already been observed in species of Lamiinae for members of family GH5_2. The first GH5_2 proteins described in species of Lamiinae were identified because of their ability to break down amorphous cellulose (Sugimura *et al.*, 2003; Wei *et al.*, 2006), which correlated with the fact that all microbial GH5_2 proteins that have been characterized so far were cellulolytic (Aspeborg *et al.*, 2012). Subsequently, by investigating the substrate specificity of six GH5_2 family members in the lamiine *Aprionа japonica*, we realized that not all six were active on amorphous cellulose – one of them evolved to be active only on xylan (Pauchet *et al.*, 2014). Interestingly, this GH5_2 xylanase processed arabinoxylan polysaccharides in a manner distinct from classical xylanase families, such as GH10 and GH11 (Pauchet *et al.*, 2020). In addition, we observed the same situation in *A. glabripennis*, another Lamiinae, where we found one GH5_2 family member – orthologous to the one in *A. japonica* – harbouring endo-β-1,4-xylanase activity, together with another GH5_2 family member having a specific activity against xyloglucan (McKenna *et al.*, 2016). Overall, these earlier studies exemplified the fact that the single CAZyme family GH5_2

evolved in species of Lamiinae in such a way that its members could process at least three PCW polysaccharides, namely amorphous cellulose, xylan and xyloglucan. Whether this holds true for GH5_2 in other subfamilies of Cerambycidae remains uncertain. A similar situation has been observed in leaf beetles (Chrysomelidae) and weevils (Curculionidae) whereby members of the typically cellulolytic CAZyme family GH45 evolved to breakdown xyloglucan instead of amorphous cellulose (Busch *et al.*, 2018; Busch *et al.*, 2019), indicating that changes in substrate specificity could have happened in many CAZyme families during the course of their evolution, in Cerambycidae, and more generally in Phytophaga beetles.

Linked to the previous point, we observe a trend that species whose genomes do not harbour genes encoding GH5_2 proteins, such as all Prioninae and most Cerambycinae, show an increased number of copies of GH45 genes. In the species we investigated, those harbouring GH5_2 proteins possess between one and three GH45 copies, whereas those species lacking GH5_2 proteins harbour up to eight GH45 copies. We speculate that some of the substrates processed by GH5_2 enzymes, such as cellulose, xyloglucan and xylan, may be broken down by GH45 enzymes in species lacking the former CAZyme family, in a kind of compensation effect.

When we plotted the distribution of PCWDE families against the species phylogeny, it became clear that the former follows the latter. In this context, it seems that the taxonomic origin of larval host plants used by these borers did not play a major role in the evolutionary history of PCWDEs in Cerambycidae. Several other aspects apart from host plant species could play a role in shaping the complement of PCWDEs in cerambycid beetles. For example, cerambycid larvae develop in host plants that vary in condition from healthy to dead, and whose tissues vary from moist to dry. They have also evolved to develop in various plant tissues (outer bark, inner bark, cambium, sapwood and heartwood), which differ in their composition of structural polysaccharides (Popper, 2008; Doblin *et al.*, 2010). These conditions and tissues of woody plants greatly vary in their nutritional quality for wood-boring insects (Hanks, 1999; Haack, 2017), and could thus influence the needs for particular classes of PCWDEs. Therefore, it is reasonable to believe that larvae of Cerambycidae feeding on decaying wood, because of the action of fungal enzymes targeting lignocellulose (Schwarze, 2007), may require a less diverse set of PCWDEs than species feeding on living woody tissues, for example, cambium. However, by looking at the distribution of PCWDE-encoding gene families in our dataset in the context of what the corresponding species use as food sources, the nutritional quality of the latter does not seem to correspond with the former. The highest diversity of PCWDE families in species of Cerambycinae, compared to other subfamilies of Cerambycidae, could be explained by a higher rate of HGT events – or one 'big' HGT event in the stem lineage of this subfamily.

Based on our results, the complement of PCWDEs harboured by Cerambycidae comprises at least ten CAZyme families. However, we anticipate that this number may increase as more species of this beetle family are studied. We conclude that Cerambycidae possess a diverse and dynamic arsenal of PCWDEs, allowing them to break down most components of the PCW, which comprises the bulk of their diet. Future research will further our understanding of the evolution of these enzyme families by functionally characterizing them, especially in terms of substrate specificity. Here, we provide the most extensive sequencing effort targeting this family of xylophagous beetles to date. We hope that these newly generated data will be widely useful to other researchers and will contribute advancing our knowledge of the fascinating physiology and ecology of the family Cerambycidae.

Author Contributions

NRS and YP conceived the study; PS, OD and SA collected the specimens used here and identified them; NRS, SGS, YO, BH, DDM and YP designed the experiments; NRS, SGS, VL and YP performed the experiments; NRS, SGS, VL, RK, BH, DDM and YP analysed the data; NRS, PS, DDM and YP wrote the paper with contributions from all co-authors. All authors read and approved the final manuscript.

Supporting Information

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

Figure S1. Alignment of GH28 from cerambycid beetles excluding Lamiinae with sequences from other Phytophaga beetles. Conserved amino acid residues are shaded and the five conserved disulphide bridges are highlighted with brackets. Asterisks and arrows indicate the positions of highly conserved putative functional important amino acid residues and catalytically active aspartates. Bold GH28s with taxon abbreviated names indicate the newly added ones from this study.

Figure S2. Amino acid alignment of glycoside hydrolase family 28 (GH28) derived from species of Lamiinae. Conserved amino acid residues are shaded and the three conserved disulfide bridges are highlighted with brackets. Asterisks and arrows indicate the positions of highly conserved putative functional important amino residues and catalytically active aspartates residues. Bold GH28s with taxon abbreviated names indicate the newly added ones from this study.

Acknowledgements

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch (MPI-CE, Jena) for technical support. We thank Thomas Kolling (Materialprüfanstalt Brandenburg GmbH; Eberswalde, Germany) for providing *Hylotrupes bajulus* specimens and Prof. John Terblanche (Centre for Invasion Biology, Stellenbosch University; South Africa) for specimens of

Cacosceles newmannii. We thank Luc Souchet and Frédéric Mouy (Office National des Forêts; France) for collecting authorizations of Cerambycid larvae in the national forests of Bellême and Orléans. This work was supported by the Max Planck Society. YP acknowledges support from the Deutsche Forschungsgemeinschaft (DFG; PA2808/4-1). DM acknowledges support from the U.S. National Science Foundation (DEB: 1355169). The authors declare no conflict of interest.

Open access funding enabled and organized by Projekt DEAL.

Data availability statement

The following supporting data are freely accessible on EDMOND – the Open Research Data Repository of the Max Planck Society with DOI: https://doi.org/10.17617/3.56.

- Supporting data 1 contains Table S1: cytochrome oxidase I (COI) barcodes extracted from the transcriptome of each species analyzed; Table S2: BUSCO analysis for each transcriptome assembly; Table S3: Assembly statistics for each transcriptome; Table S4: Distribution of CAZyme families in each transcriptome assembly; Table S5: Accessions of the raw data and assembled transcriptomes deposited in the corresponding NCBI databases (ncbi_sra and ncbi_tsa); Table S6: List of the primers used for RACE-PCRs.
- Supporting data 2 'Species phylogeny' contains the codon-based nucleotide alignment (FASTA file), the IQ-TREE log file and the resulting tree file (Newick file) used to generate Fig. 1.
- Supporting data 3 contains all the nucleotide sequences (open reading frames) for the 340 putative PCWDEs identified in cerambycid transcriptomes.
- Supporting data 4 contains the sequence alignments (FASTA), IQ-TREE log files and resulting tree files (Newick) corresponding to the phylogenetic analyses reported in Fig. 3.
- Supporting data 5 contains the draft genome assemblies generated from the *Phymatodes lengi* and *Xylotrechus colonus* genome sequence data retrieved from Genbank.

References

Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H. 3rd & Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evolutionary Biology*, **12**, 186. https://doi.org/10.1186/1471-2148-12-186.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Buchner, P. (1928) Holƶnahrung und symbiose. *Holƶnahrung und Symbiose: Vortrag Gehalten auf dem X. Internationalen Ƶloologentag Ƶu Budapest am 8 September 1927*, pp. 5–64. Springer Berlin Heidelberg, Berlin, Heidelberg.

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**, 633–650. https://doi.org/10.1111/imb.12500.

Busch, A., Danchin, E.G.J. & Pauchet, Y. (2019) Functional diversification of horizontally acquired glycoside hydrolase family 45 (GH45) proteins in Phytophaga beetles. *BMC Evolutionary Biology*, **19**, 100. https://doi.org/10.1186/s12862-019-1429-9.

Capella-Gutierrez, S., Silla-Martinez, J.M. & Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Chararas, C. & Chipoulet, J.-M. (1982) Purification by chromatography and properties of a β-glucosidase from the larvae of *Phoracantha semipunctata*. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, **72**, 559–564. https://doi.org/10.1016/0305-0491(82)90506-5.

Chararas, C. & Chipoulet, J.-M. (1983) Studies on the digestion of cellulose by the larvae of the eucalyptus borer, *Phoracantha semipunctata* (Coleoptera: Cerambycidae). *Australian Journal of Biological Sciences*, **36**, 223–234. https://doi.org/10.1071/BI9830223.

Chipoulet, J.-M. & Chararas, C. (1985a) Purification and partial characterization of a cellobiase from the larvae of *Rhagium inquisitor*. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, **82**, 327–332. https://doi.org/10.1016/0305-0491(85)90249-4.

Chipoulet, J.-M. & Chararas, C. (1985b) Survey and electrophoretical separation of the glycosidases of *Rhagium inquisitor* (Coleoptera: Cerambycidae) larvae. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, **80**, 241–246. https://doi.org/10.1016/0305-0491(85)90203-2.

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

Foglietti, M., Courtois, J., Chararas, C. & Levaditou, V. (1971) Pectinases from xylophagous insect *Rhagium inquisitor* L. *Comptes Rendus des Séances de la Société de Biologie et de Ses Filiales*, **165**, 1019–1022.

Grabherr, M.G., Haas, B.J., Yassour, M. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652. https://doi.org/10.1038/nbt.1883.

Haack, R.A. (2017) Feeding biology of cerambycids. *Cerambycidae of the World: Biology and Pest Management* (ed. by Q. Wang), pp. 105–132. CRC Press, Boca Raton, Florida.

Haddad, S., Shin, S., Lemmon, A.R. *et al.* (2018) Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). *Systematic Entomology*, **43**, 68–89. https://doi.org/10.1111/syen.12257.

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

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518–522. https://doi.org/10.1093/molbev/msx281.

Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. (2010) CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Kapustin, Y., Souvorov, A., Tatusova, T. & Lipman, D. (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. *Biology Direct*, **3**, 20. https://doi.org/10.1186/1745-6150-3-20.

Kirsch, R., Gramzow, L., Theissen, G., Siegfried, B.D., Ffrench-Constant, R.H., Heckel, D.G. & 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 Biochemistry and Molecular Biology*, **52**, 33–50. https://doi.org/10.1016/j.ibmb.2014.06.008.

Kukor, J.J. & Martin, M.M. (1986) Cellulose digestion in *Monochamus marmorator* Kby. (Coleoptera: Cerambycidae): role of acquired fungal enzymes. *Journal of Chemical Ecology*, **12**, 1057–1070. https://doi.org/10.1007/BF01638996.

Kukor, J.J., Cowan, D.P. & Martin, M.M. (1988) The role of ingested fungal enzymes in cellulose digestion in the larvae of cerambycid beetles. *Physiological Zoology*, **61**, 364–371.

Lee, S.J., Kim, S.R., Yoon, H.J. *et al.* (2004) cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. *Comparative Biochemistry and Physiology B – Biochemistry & Molecular Biology*, **139**, 107–116. https://doi.org/10.1016/j.cbpc.2004.06.015.

Lee, S.J., Lee, K.S., Kim, S.R. *et al.* (2005) A novel cellulase gene from the mulberry longicorn beetle, *Apriona germari*: gene structure, expression, and enzymatic activity. *Comparative Biochemistry and Physiology B – Biochemistry & Molecular Biology*, **140**, 551–560. https://doi.org/10.1016/j.cbpc.2004.12.003.

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

Liu, J., Song, K., Teng, H., Zhang, B., Li, W., Xue, H. & Yang, X. (2015) Endogenous cellulolytic enzyme systems in the longhorn beetle *Mesosa myops* (Insecta: Coleoptera) studied by transcriptomic analysis. *Acta Biochimica et Biophysica Sinica (Shanghai)*, **47**, 741–748. https://doi.org/10.1093/abbs/gmv070.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. & Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**, D490–D495. https://doi.org/10.1093/nar/gkt1178.

Mansour, K. & Mansour-Bek, J.J. (1934) The digestion of wood by insects and the supposed role of micro-organisms. *Biological Reviews*, **9**, 363–382. https://doi.org/10.1111/j.1469-185X.1934.tb01252.x.

Martin, M.M. (1983) Cellulose digestion in insects. *Comparative Biochemistry and Physiology part A-Physiology*, **75**, 313–324. https://doi.org/10.1016/0300-9629(83)90088-9.

Martin, M.M. (1991) The evolution of cellulose digestion in insects. *Philosophical Transactions of the Royal Society B – Biological Sciences*, **333**, 281–288. https://doi.org/10.1098/rstb.1991.0078.

Marvaldi, A.E., Duckett, C.N., Kjer, K.M. & Gillespie, J.J. (2009) Structural alignment of 18S and 28S rDNA sequences provides insights into phylogeny of Phytophaga (Coleoptera: Curculionoidea and Chrysomeloidea). *Zoologica Scripta*, **38**, 63–77. https://doi.org/10.1111/j.1463-6409.2008.00360.x.

Mattson, W.J. Jr (1980) Herbivory in relation to plant nitrogen content. *Annual Review of Ecology and Systematics*, **11**, 119–161. https://doi.org/10.1146/annurev.es.11.110180.001003.

McKenna, D.D., Scully, E.D., Pauchet, Y. *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**, 227. https://doi.org/10.1186/s13059-016-1088-8.

McKenna, D.D., Shin, S., Ahrens, D. *et al.* (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 24729–24737. https://doi.org/10.1073/pnas.1909655116.

Mei, H.Z., Xia, D.G., Zhao, Q.L., Zhang, G.Z., Qiu, Z.Y., Qian, P. & Lu, C. (2016) Molecular cloning, expression, purification and characterization of a novel cellulase gene (Bh-EGaseI) in the beetle *Batocera horsfieldi*. *Gene*, **576**, 45–51. https://doi.org/10.1016/j.gene.2015.09.057.

Mewis, K., Lenfant, N., Lombard, V. & Henrissat, B. (2016) Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Applied and Environmental Microbiology*, **82**, 1686–1692. https://doi.org/10.1128/AEM.03453-15.

Misof, B., Liu, S., Meusemann, K. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767. https://doi.org/10.1126/science.1257570.

Monné, M.L., Monné, M.A. & Wang, Q. (2017) General morphology, classification, and biology of Cerambycidae. *Cerambycidae of the World: Biology and Pest Management* (ed. by Q. Wang), pp. 1–70. CRC Press, Boca Raton, Florida.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274. https://doi.org/10.1093/molbev/msu300.

Nie, R., Vogler, A.P., Yang, X.K. & Lin, M. (2021) Higher-level phylogeny of longhorn beetles (Coleoptera: Chrysomeloidea) inferred from mitochondrial genomes. *Systematic Entomology*, **46**, 56–70. https://doi.org/10.1111/syen.12447.

Oberprieler, R.G., Marvaldi, A.E. & Anderson, R.S. (2007) Weevils, weevils, weevils everywhere. *Zootaxa*, **1668**, 491–520.

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. 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**, 1517–1525. https://doi.org/10.1002/cbic.201900687.

Pearson, W.R., Wood, T., Zhang, Z. & Miller, W. (1997) Comparison of DNA sequences with protein sequences. *Genomics*, **46**, 24–36. https://doi.org/10.1006/geno.1997.4995.

Peters, R.S., Krogmann, L., Mayer, C. *et al.* (2017) Evolutionary history of the hymenoptera. *Current Biology*, **27**, 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027.

Popper, Z.A. (2008) Evolution and diversity of green plant cell walls. *Current Opinion in Plant Biology*, **11**, 286–292. https://doi.org/10.1016/j.pbi.2008.02.012.

Schwarze, F.W.M.R. (2007) Wood decay under the microscope. *Fungal Biology Reviews*, **21**, 133–170. https://doi.org/10.1016/j.fbr.2007.09.001.

Scrivener, A., Watanabe, H. & Noda, H. (1997) Diet and carbohydrate digestion in the yellow-spotted longicorn beetle *Psacothea hilaris*. *Journal of Insect Physiology*, **43**, 1039–1052. https://doi.org/10.1016/S0022-1910(97)00063-2.

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**, 850. https://doi.org/10.1186/1471-2164-14-850.

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, *Psacothea hilaris*. *European Journal of Biochemistry*, **270**, 3455–3460. https://doi.org/10.1046/j.1432-1033.2003.03735.x.

Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–W612. https://doi.org/10.1093/nar/gkl315.

Svacha, P. & Lawrence, J.F. (2014) Cerambycidae Latreille, 1802. *Handbook of Zoology. Arthropoda: Insecta. Coleoptera, Beetles:*

Morphology and Systematics (Phytophaga), Vol. **3** (ed. by R.G. Beutel and R.A.B. Leschen), pp. 77–178. Walter de Gruyter, Berlin.

Waterhouse, R.M., Seppey, M., Simao, F.A. *et al.* (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, **35**, 543–548. https://doi.org/10.1093/molbev/msx319.

Wei, Y.D., Lee, K.S., Gui, Z.Z. *et al.* (2006) Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, *Apriona germari*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, **145**, 220–229. https://doi.org/10.1016/j.cbpb.2006.07.007.

Accepted 21 April 2021
First published online 11 May 2021