Diversity, evolutionary history and functional characterization of plant cell wall degrading enzymes in beetles of the family Cerambycidae

Dissertation

To Fulfill the

Requirements for the Degree of

“doctor rerum naturalium” (Dr. rer. nat.)

Submitted to the Council of the Faculty of

Biological Sciences

of Friedrich Schiller University Jena

by M.Sc. Na Ra Shin

born on 15.05.1988 in Gwangju, South Korea

Reviewer:

1. Prof. Dr. David G. Heckel, Max-Planck-Institut für chemische Ökologie, Jena
2. Prof. Dr. Ralf Oelmüller, Friedrich-Schiller-Universität Jena
3. Dr. Jean-Guy Berrin, INRAE, Fungal Biodiversity and Biotechnology (BBF), Marseille, France

Date of public defense: 23.02.2023

I. Table of Contents

II. List of Abbreviations

AA	auxiliary activities
BUSCO	Benchmarking Universal Single-Copy Orthologs
CE	carbohydrate esterase
CRISPR	clustered regularly interspaced short palindromic repeats
EC	enzyme commission
gDNA	genomic DNA
GH	glycoside hydrolase
GH10	glycoside hydrolase family 10
GH11	glycoside hydrolase family 11
GH28	glycoside hydrolase family 28
GH43_26	glycoside hydrolase family 43 subfamily 26
GH45	glycoside hydrolase family 45
GH48	glycoside hydrolase family 48
GH5_2	glycoside hydrolase family 5 subfamily 2
GH5_8	glycoside hydrolase family 5 subfamily 8
GH53	glycoside hydrolase family 53
GH7	glycoside hydrolase family 7
GH9	glycoside hydrolase family 9
GT	glycosyltransferases
HGT	horizontal gene transfer
ORF	open reading frame
PCW	plant cell wall
PCWDEs	plant cell wall degrading enzymes
PG	polygalaturonase
PL	polysaccharide lyases
RNAi	RNA interference

III. List of Figures

They are in you and me; they created us, body and mind;

and their preservation is the ultimate rationale for our existence.

They have come a long way, those replicators.

Now they go by the name of genes, and we are their survival machines.

"The selfish gene", Richard Dawkins

1. General Introduction

1.1 Overview of genetic studies in history

In history, the first document on an entomological observation of insect body structure and its behavior was published by Aristotle (384 to 322 BC) (Essig 1936; Reynolds 2019). Although many inaccurate studies about insects came out before the Renaissance, the invention of the microscope by Antony van Leeuwenhoek (1632 – 1723) led to new insight into the understanding of insect morphology. The microscope also allowed researchers to observe cells and chromosomes (Hooke 1665, Waldeyer 1888). Moreover, new theories about the concept of evolution were developed in 1858: Darwin proposed the descent of species from a common ancestor and the concept of natural selection which drives evolution (Darwin 1868). Gregor Mendel advanced the laws of inheritance in 1865 with the first introduction of the transmission of heritable traits (Abbott and Fairbanks 2016; Mendel 1866). These breakthrough theories on the interconnectedness of all organisms and genetic inheritance became the foundation of genetics (Koonin 2009). The first genetic map of fruit fly (*Drosophila melanogaster*) genes was attained in 1911 (Sturtevant 1913), but the foundational chemistry of the genetic materials for inheritance (such as DNA) still remained elusive. Only in the early 1950s, when the molecular structure and information of DNA was identified using x-ray diffraction by Watson and Crick (Watson and Crick 1953), did modern molecular genetics develop. Since the discovery of the double-helix structured DNA, scientists have succeeded in isolating DNA from various species and demonstrating DNA transcription and translation in living organisms. In 1977, technological developments from Frederick Sanger allowed the generation of the first complete genome of a bacteriophage (Sanger et al. 1977). Earlier research in genetics revealed foundational theories into genomic elements as well as transposable elements which expanded knowledge of evolutionary relationships (Biémont 2010; Robertson 1993). However, a new era in genetics dawned with the advent of genome sequencing. In 1999, the Human Genome Project began as an international cooperative work (Lander et al. 2001). Over 13 years, this project uncovered the whole human genome sequence. In 2000, the first insect genome sequence for *Drosophila melanogaster* was published (Adams et al. 2000; Lander et al. 2001).

Understanding and applications in molecular biology have developed with the advancement of sequencing technologies. In the early stages of DNA sequencing, single gene markers - such as those encoding ribosomal RNA -or fully sequenced mitochondrial genomes allowed for comparative analyses between

species as well as phylogenetic analyses to study evolutionary relationships (Cameron et al. 2009; Lin et al. 2002; Sheffield et al. 2008). However, individual gene sequence are not sufficient to infer accurate species phylogenies due to gene duplication or horizontal gene transfer events, and mitochondrial genomes are not adequate to provide the view of species divergences (Rokas et al. 2003; Song et al. 2010; Trautwein et al. 2012). Furthermore, early studies with sequencing data focused on isolating and characterizing specific genes related to certain tissues or in specific developmental stages (Ferreira et al. 2001; Girard and Jouanin 1999a; Sugimura et al. 2003). Twenty years after the rise of next-generation sequencing, molecular biology technologies have developed further, allowing for more in-depth understanding of genes or species of interest. Now, geneticists can employ orthologous gene sets generated from accumulated genome sequencing data extracted from many species (Haddad et al. 2018; Johnson et al. 2018; McKenna et al. 2019). Comparative analyses are not limited to specific genes or mitochondrial genomes but are expanded to the chromosome level. Today, we can study evolutionary relationships between species, as well as the evolution of significant gene families, using cheaper and more efficient methods.

1.2 The family Cerambycidae

Cerambycidae, also known as longhorned beetles or longicorns, are wood-boring (xylophagous) insects within the Phytophaga (plant-feeding) clade of beetles. Phytophaga comprises two sister superfamilies: Chrysomeloidea and Curculionoidea (Coleoptera). Chrysomeloidea includes longhorned beetles and leaf beetles, while Curculionoidea includes weevils, snout beetles, and bark beetles. Cerambycid beetles account for a large fraction of the entire beetle order (estimated 5,300 genera and 36,300 extant species out of 350,000 described coleopteran species) (Linsley 1959; Švácha and Lawrence 2014). They present the highest species-richness and abundance in tropical regions, but are distributed worldwide. To date, eight subfamilies are known in Cerambycidae; Lamiinae, Cerambycinae, Spondylidinae, Lepturinae, Prioninae, Dorcasominae, Parandrinae, and Necydalinae (Haddad et al. 2018) (Figure 1). In the eight subfamilies,, Lamiinae is the largest subfamily with around 21,000 species, and the next species-rich subfamily is Cerambycinae in this family (Monné and Wang 2017). Both are relatively large groups compared to other cerambycid subfamilies, which are characterized by low species-richness and a limited distribution (Švácha and Lawrence 2014).

Within the life cycle of cerambycid beetles, larvae are predominantly associated with wood to acquire nutrients as they bore deep into woody tissue (Hanks 1999; Linsley 1959). Most cerambycid larvae take six months to several years to develop inside the host plant into the pre-adult life cycle. Species of Lamiinae have the shortest development period among cerambycid beetles. However, Lamiinae also exhibits different developmental periods depending on dietary conditions (Linsley and Chemsak 1961). They showed a developmental delay of several years when grown with dry and processed wood, compared to three to six months of developmental periods when larvae feed on stems of plants. Because these beetles break down the decomposed trees and break polysaccharide down to monomers, longhorned beetles are considered significant players in forest ecosystems (Haack 2017). However, these beetles also cause damage and disease in economically important orchards and wild trees. One of the invasive pests in cerambycid beetles is *Anoplophora glabripennis*, which causes large economic losses for the Asian forestry industry and ecosystem (Meng et al. 2015). Another invasive pest species, *Monochamus* (Lamiinae, Cerambycidae) is a severe threat to the forestry industry world-wide, especially in Asia and Europe, because it feeds on healthy conifers and spread nematodes, causing pine wilt disease (Akbulut and Cardak 2012).

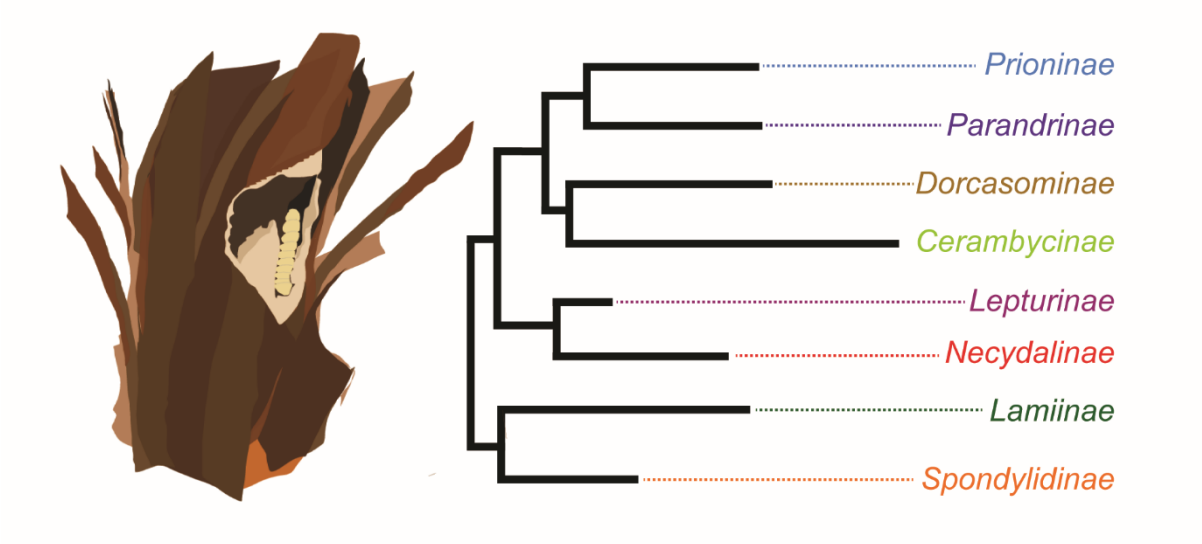

Figure 1. Schematic view of the phylogenetic relationships between cerambycid subfamilies. *Cerambycid beetles are recognized as Phytophaga and develop inside decomposed trees in the larval stage as xylophagous beetles. Cerambycidae comprises eight known subfamilies;Prioninae, Parandrinae, Dorcasominae, Cerambycinae, Lepturinae, Necydalinae, Lamiinae, and Spondylidinae. Different color codes represent the subfamilies of cerambycid beetles (This figure was modified by Haddad et al (2018)).*

1.3 Plant cell walls

The plant cell wall is one of the most abundant natural resources on Earth and the fundamental unit of wood structure. Plant cell walls play important supporting roles in structural maintenance and defense systems against biotic and abiotic stress (Somerville et al. 2004). Plant cell walls are made up of polysaccharides such as cellulose microfibrils, hemicellulose, pectin, and lignin with small amounts of protein, such as hydroxyproline-rich glycoproteins (Burton et al. 2010; Doblin et al. 2010; Vorwerk et al. 2004; Zhao et al. 2019). In most higher plants, cell walls are comprised of two layers, primary and secondary cell walls, which harbor different polymer components and have different mechanical and structural functions (Bacic et al. 1988) (Figure 2A). The primary cell wall is synthesized during plant growth and is mainly composed of cellulose, xyloglucan, and pectin (O'neill et al. 1990; Pauly and Keegstra 2016; Ridley et al. 2001) (Figure 2B). Due to the high proportions of xyloglucan and pectin they contain, primary cell walls are more flexible and hydrated structures than secondary cell walls (Cosgrove and Jarvis 2012; Fry 2004). In general, pectic polysaccharides account for the most significant proportion of the primary cell wall (30-50%) on a dry weight basis, followed by cellulose (15-40%), xyloglucan (20-30%), and minor amounts of structural proteins (Cosgrove and Jarvis 2012). In contrast, secondary cell walls are synthesized after cell growth completely ceased (Barnes and Anderson 2018) (Figure 2C). They are also made up of polysaccharides but are more rigid than primary cell walls due to the presence of lignin (Vanholme et al. 2010). The main constituents of secondary cell walls are also cellulose, mannan, and xylan (Mellerowicz and Sundberg 2008; Scheller and Ulvskov 2010). Woody tissues have been historically considered -nutrient-poor resources, with low amounts of the proteins, lipids, and vitamins necessary for insect development (Mattson Jr 1980). Moreover, plant cell wall polysaccharides provide a complex configuration, which is recalcitrant to degradation (Himmel et al. 2007). Although historical insight did not consider woody tissues as providing enough nutritional advantages to animals, many organisms, including cerambycid beetles, have evolved to obtain nutrients from degrading plant cell wall polysaccharides (Calderón-Cortés et al. 2012; Castillo-González et al. 2014; Linsley 1959; Weimer 1992). The following sections detail the various plant cell wall polysaccharides such as cellulose, hemicellulose (xyloglucan, xylan, and mannan), pectin, and lignin.

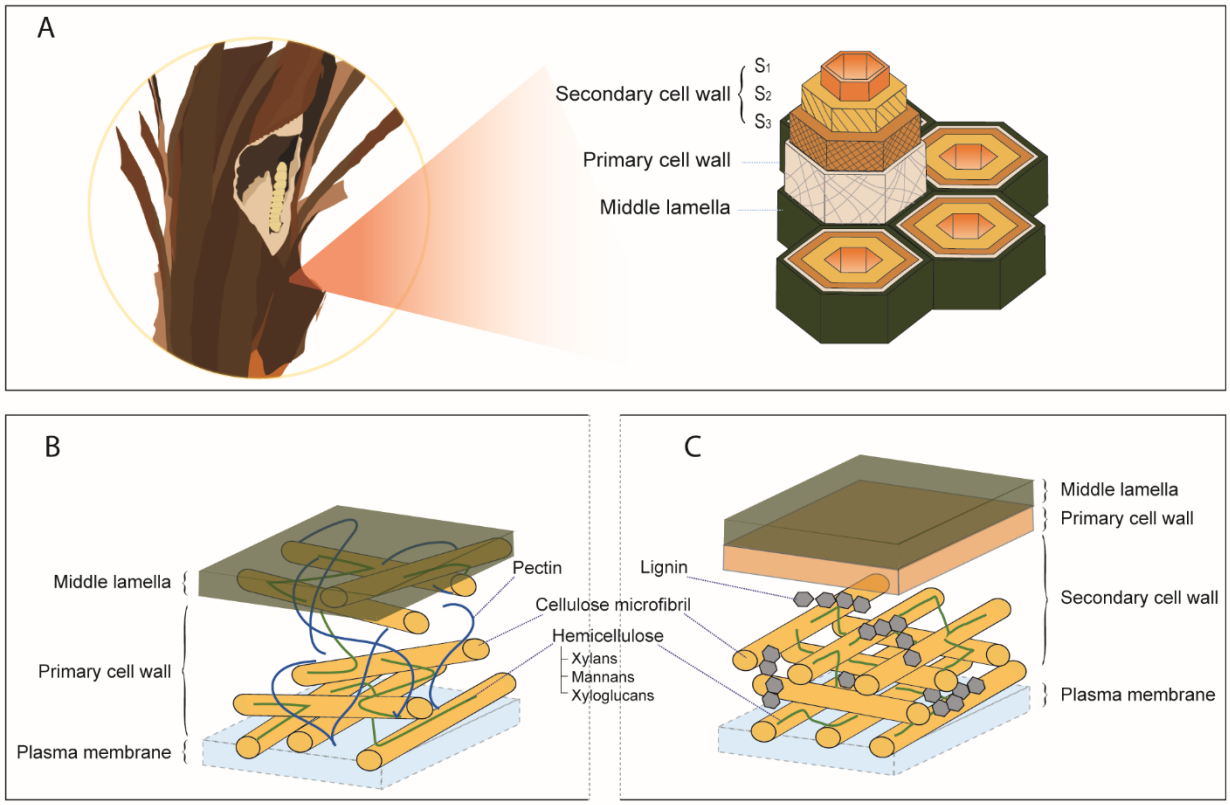

Figure 2. Schematic structure of the plant cell wall with polysaccharide components. *Plant cells are the basic unit of wood and are comprised of two cell wall layers: primary and secondary. The middle lamella forms the connection between the primary cell walls of adjacent cells. The secondary cell wall comprises three layers (S1, S2, and S3) (Figure 2A). The main component of both primary and secondary cell walls is cellulose microfibril. Hemicellulose corresponding to xylans, mannans, and xyloglucans are also found commonly in both cell walls. However, pectin is identified in the primary cell wall (Figure 2B), whereas it may be absent in the secondary cell wall. Lignin is found in the secondary cell wall as one of the major components but not in the primary cell wall (Figure 2C). (This figure was modified from Zhao (Zhao et al 2019))*

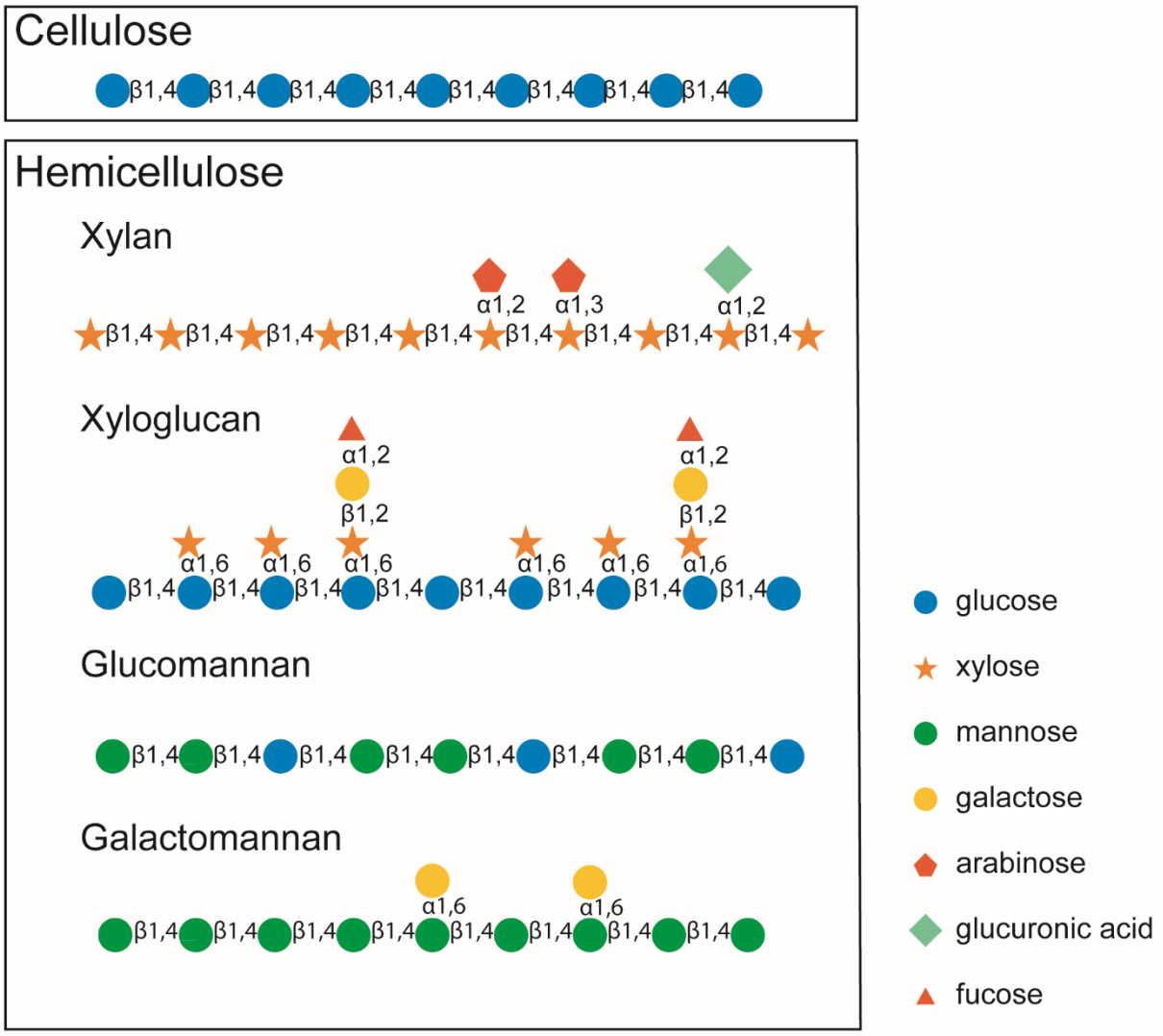

Figure 3. Schematic structure of cellulose and hemicelluloses in plant cell wall polysaccharides. *Plant cell walls contain the polysaccharides, cellulose, hemicellulose, and pectin. Cellulose is a linear polymer linked by β-1,4 linkages between glucose units. Hemicelluloses (xylan, xyloglucan, glucomannan, and galactomannan) are heterogeneous polysaccharides constituting a β-1,4-linked backbone - made up of different sugar units such as xylose, mannose, glucose, and galactose - with side chains (This figure was modified by Martens (Martens et al. 2011))*

1.3.1 **Cellulose, hemicellulose, and pectin**

Cellulose is the most abundant polysaccharide on Earth and a linear polymer composed of 1,4-linked β-D-glucose residues (Figure 3). Cellulose organizes into two different types of structure - crystalline cellulose or amorphous cellulose - depending on the arrangement of hydrogen bonds (Hearle and Peters 2013). While crystalline cellulose is formed in long, regularly-arranged polymer chains by strong hydrogen bonds, amorphous cellulose is a twisted polymer formed by weak hydrogen bonds (Watanabe and Tokuda 2010). Despite accumulated studies on the impact of the different ratios of crystalline versus amorphous cellulose

on the mechanical properties of cellulose, the distribution of both type of cellulose in the primary and secondary cell walls is still not well-characterized (Cosgrove and Jarvis 2012; Rongpipi et al. 2019).

Hemicellulose comprises non-cellulosic polysaccharide components of plant cell walls sharing a β-1,4-linked backbone with neutral sugars such as glucose, xylose, or mannose with short side chains (Scheller and Ulvskov 2010). Hemicellulose polysaccharides like xyloglucan, xylan, and gluco- and galactomannan play important roles in stabilizing the structure of plant cell walls by interacting with cellulose microfibrils and lignin (Figure 3). In particular, xyloglucan binds to cellulose microfibrils in the primary cell wall, which are important for stabilizing cell wall structure (Burton et al. 2010; Cosgrove and Jarvis 2012; Scheller and Ulvskov 2010). Recent studies found that hydrogen bonding plays an important role in the interaction between xylan and cellulose microfibrils (Donev et al. 2018; Heinonen et al. 2022). Once xylan adsorbs to cellulose, hydrogen bonds are formed between cellulose and xylan, which extend cellulose microfibrils. This formation can impact the rigidity and flexibility of plant cell walls. Many observations on the binding and interactions between hemicellulose and cellulose have been published (Berglund et al. 2016; Heinonen et al. 2022; Park and Cosgrove 2015).

Most pectins are structurally complex polymers sharing a α-1,4-linked galacturonic acid (GalA) backbone (Figure 4). The pectic polymers are classified into four subclasses: homogalacturonan (HG), rhamnogalacturonan I and II (RG-I, II), and xylogalacturonan (XGA). Both HG and XGA consist of a GalA backbone with esterified/acetylated side-chains, while the the GalA backbone of XGA features xylose side chains (Thibault and Ralet 2008). The backbone of both RGs also includes GalA residues. RG-I contains a backbone of the disaccharide repeat unit (2)- α-L-Rha(1,4)- α-D-GalA(1-) with linear α-l-arabinofuranosyl and β-d-galactopyranosyl side chains. On the other hand, RG-II consists of a GalA backbone decorated with side branches as a monosaccharide, octasaccharide, and two structurally different disaccharides. Pectins modulate the hydration of the primary cell wall, which is important for cell wall expansion and cell separation during plant development (Wolf et al. 2009). Although the amount of pectin found in the secondary cell wall is very small, many studies have addressed its importance in the secondary cell wall biosynthesis and modification (Xiao and Anderson 2013).

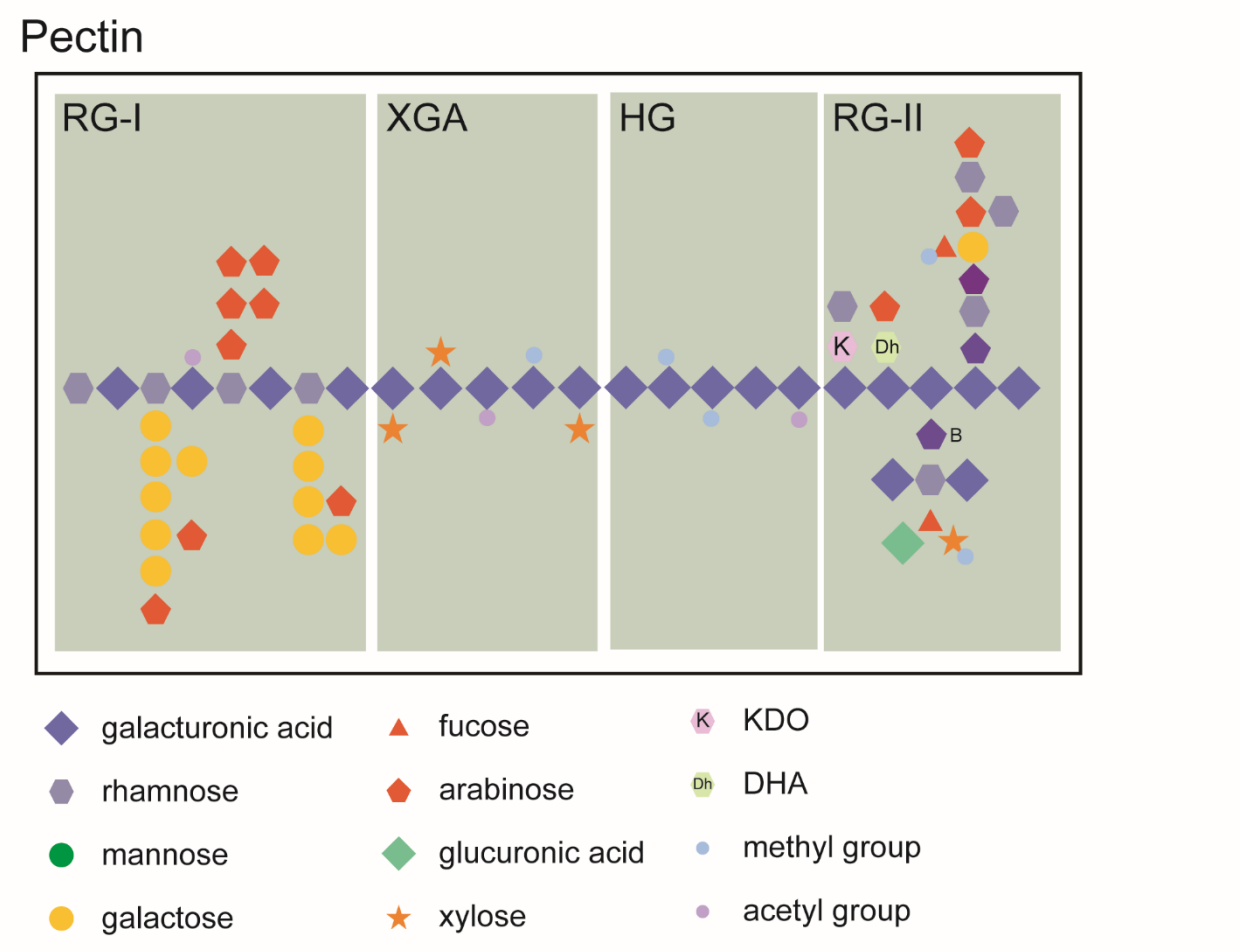

Figure 4. Schematic structure of pectin in plant cell wall polysaccharides. *Pectin is a heteropolysaccharide predominantly containing galacturonic acid residues. Pectins can be divided into four subclasses with different side chains: homogalacturonan (HG), rhamnogalacturonan I and II (RG-I, II), and xylogalacturonan (XGA) (This figure was modified by (Mohnen 2008)).*

1.3.2 Lignin

Lignin is one of the most abundant polymers in wood and is mainly found in the secondary cell wall of the plant (Boudet 1998). The network of cellulose microfibrils and hemicelluloses is embedded in lignin, which promotes the integrity, rigidity, and hydrophobicity of the secondary cell wall (Cosgrove and Jarvis 2012; Kang et al. 2019). Also, the lignified structure enables the plant to transport water and nutrients through plant tissues for plant growth, as it strengthens the woody tissue and reduces permeability (Vanholme et al. 2019; Zhao 2016). Lignin is composed of alkyl-aromatic polymers as a polyphenolic material. The structure and contents of lignin are very dynamic between plant species, tissues, environmental conditions, and even developmental stages (Vanholme et al. 2010). Lignin includes three cell-specific formations corresponding to p-hydroxyphenyl (1H, H-units), guaiacyl (1G, G-units), and

syringyl (1S, S-units) structures. For example, while G-units are mainly found in softwood lignin, both G- and S-units are abundant in the hardwood lignin (Katahira et al. 2018).

1.4 Carbohydrate active enzymes (CAZymes)

Carbohydrates are one of the large groups of organic molecules on Earth and important components in all living organisms (Stern and Jedrzejas 2008). Among them, polysaccharides are used as an energy source to provide sufficient nutrients for the organism. It is also the most abundant and essential organic component in plant structure (Doblin et al. 2010; Stern and Jedrzejas 2008). Carbohydrate is a diverse group of organic compound with different chemical and physical properties such as solubility, crystallization state, and aggregated structure in plant cell walls (Cummings and Stephen 2007). Carbohydrate active enzymes (CAZymes) are enzymes related to the synthesis and degradation of diverse carbohydrates (Drula et al. 2021). Since the first classification of cellulases into distinct enzyme families based on amino acid similarities (Henrissat et al. 1989), the CAZy database (http://www.cazy.org) has classified CAZymes into the following classes based mainly on sequence similarities: glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and those with auxiliary activities (AA) (Drula et al. 2021). The following sections describe the different enzyme classes in detail.

1.4.1 Glycoside hydrolases

Glycoside hydrolases are common enzymes in almost all organisms and constitute the largest group of CAZymes. These enzymes are currently classified into 173 families based on sequence similarities and folds of proteins (Drula et al. 2021). The number of GH families has been increased on a regular basis because new enzymes which do not belong to previously classified families are continuously discovered. GH enzymes catalyze the hydrolysis of glycosidic O-, N-, and S-linkage bonds mediated by two carboxyl groups. Depending on the cleavage regions on the substrate, they are divided into two types: endo- and exo-acting enzymes. All cellulase is classified into glycoside hydrolase families (Davies and Henrissat 1995). Most cellulases are endo-acting enzymes, which cleave a polymer substrate within the middle of the chain. Exo-acting enzymes act on the non-reducing and/or reducing end of the polymer substrate chain. GH enzyme families are also categorized into either retaining or inverting enzymes (Koshland Jr 1953) (Figure 5). Both types of enzyme possess two carboxyl groups that act as an acid (proton donor) and a base (nucleophile assistance). Retaining enzymes hydrolyse polymers via a double displacement reaction

(Figure 5A), whereas inverting enzymes hydrolyse polymers via a single displacement reaction (Figure 5B). Both mechanisms include oxocarbenium ion-like transition states. However, the double displacement reaction involves a covalent glycosyl-enzyme intermediate, which causes the conformation to be retained in the catalytic reaction.

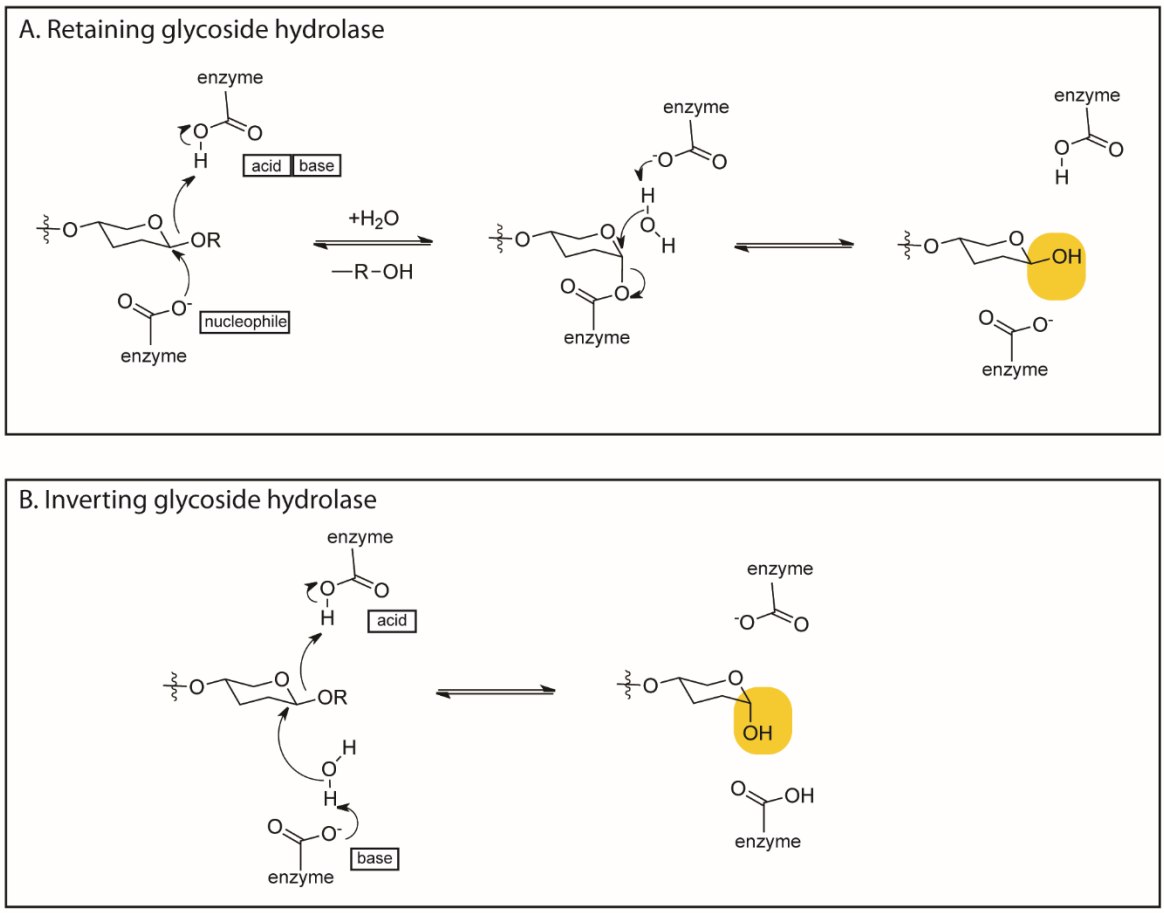

Figure 5. Schematic illustrating the catalytic activity of glycoside hydrolases. *Glycoside hydrolases catalyze either retaining or inverting reactions for hydrolysis. Both steps utilize an acid/base from two amino acid side chains, such as glutamate and aspartate. In the retaining reaction (Figure 5A), acid/base catalyst carboxyl groups provide a hydrogen ion to the C terminus at the cleavage site, and a hydroxyl ion is generated from a water molecule for the C1 terminus. The newly generated C terminus remains in the original conformation as the result of retaining glycoside hydrolase. In the inverting reaction (Figure 5B), the acid catalyst group provides hydrogen nuclei to the C terminus of the cleavage site, and the base catalyst gives a hydroxyl ion from a water molecule to the C1 terminus. The final conformation of the newly generated C terminus from the inverting glycoside hydrolase is converted from the original conformation.*

1.4.2 Glycosyltransferases

Glycosyltransferases (GTs) catalyze the synthesis of the glycoside linkage using activated sugar phosphates as glycosyl donors by retention or inversion of anomeric configuration (Lairson et al. 2008). There are two types of glycosyltransferase, determined by dependence on sugar nucleotides and different folding formations in protein structure; Leloir glycosyltransferases and non-Leloir glycosyltransferases. Leloir glycosyltransferases are the sugar nucleotide-dependent enzymes, which catalyze the connection of sugar mono- or diphosphonucleotides with glycoside linkages. In contrast, non-Leloir glycosyltransferases use non-nucleotide substrates such as polyprenol pyrophosphates, polyprenol phosphate, and sugar 1-phosphates with divalent metals (Mn^{2+} and Mg^{2+}) as their substrates.

1.4.3 Polysaccharide lyases, Carbohydrate esterases, and Auxiliary activities

Polysaccharide lyases (PL) cleave uronic acid-containing polysaccharide chains using elimination mechanisms. These enzyme families are also classified according to sequence similarity and specific folding types. The PL families have different substrate preferences and produce different products (Biely 2012; Garron and Cygler 2010). Carbohydrate esterases (CE) cleave -O or -N ester bonds to modify the structure of polysaccharides for the actions of glycoside hydrolases following esterase activity. Esters are composed of acid and alcohol, so CE hydrolyzes the linkage of both pectin methyl esters and acetylated xylan. (Biely 2012). Auxiliary activities (AA) are another class of enzymes, recently classified also based on sequence similarity and protein structural information (Levasseur et al. 2013). AA includes two different groups, with nine families of ligninolytic enzymes and seven families of lytic polysaccharide mono-oxygenases. AA contains predominantly oxidative ligninolytic enzyme families for lignin depolymerization, but lytic polysaccharide monooxygenases (LPMO) also belong to this class of enzymes. Enzymes belonging to this class break down lignin as well as degrade cellulose and chitin (Hemsworth et al. 2014).

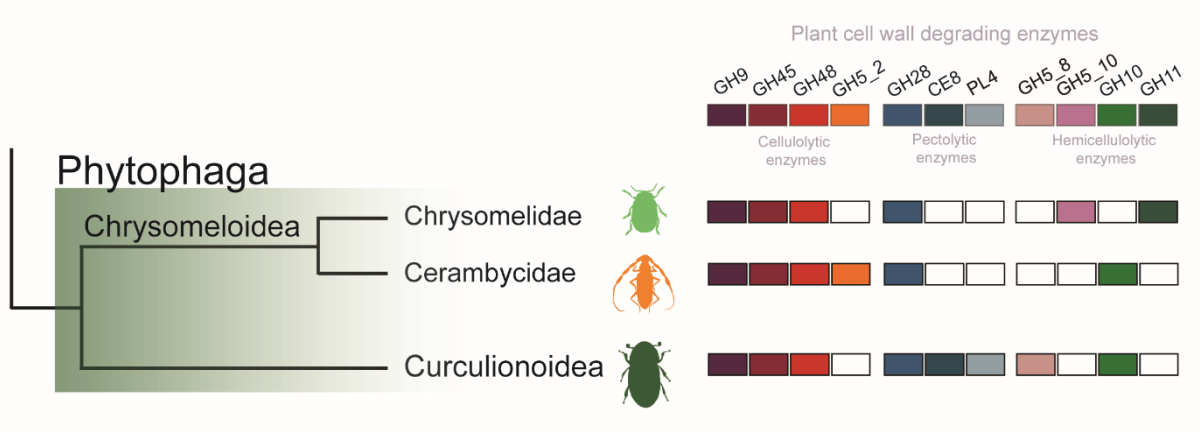

Figure 6. Schematic view of the distribution of PCWDEs in phytophagous beetles. *Phytophagous beetles constitute two sister superfamilies, Chrysomeloidea and Curculionoidea. Endogenous PCWDEs have been identified from both superfamilies and are essential enzymes for breaking down plant cell walls as their primary nutrient sources. Some PCWDEs like GH9, GH45, GH48, and GH28 seem to be conserved across the phytophagous beetle clade, while others, such as GH5_2, CE8, PL4, and most hemicellulolytic enzymes, are restricted to certain groups (This figure was modified by McKenna et al (2019))*

1.5 Plant cell wall degrading enzymes derived from insects

A subgroup of CAZymes, plant cell wall degrading enzymes are essential to break down plant cell wall polysaccharides. PCWDEs degrade most polysaccharide components, including cellulose, hemicellulose (xylan, xyloglucan, and mannans), and pectin. PCWDEs are classified into three categories based on their enzymatic activities, cellulolytic enzymes, hemicellulolytic enzymes, and pectolytic enzymes. Although many phytophagous animals use PCWDEs obtained from microsymbionts, several cases of enzymes secreted from insect digestive tracts have been reported (Girard and Jouanin 1999a; Lee et al. 2004; Pauchet and Heckel 2013; Pauchet et al. 2010; Watanabe et al. 1998; Watanabe and Tokuda 2001). The first observation of endogenous PCWDEs' presence in the gut fluid of cerambycid larvae was in the early 1900s (Linsley 1959). It contrasted the previous significant studies that cerambycid beetles digest woody tissues with cellulase generated from a symbiont yeast, which was found in the epithelial cells or a mycetoma of the insects' gut (Heitz 1927; Kukor et al. 1988; Schomann 1937). Nevertheless, undeniable evidence was published to challenge the notion that digestive enzymes originated from the symbiont. First, the difference in degraded cellulose based on the presence or absence of yeast symbionts was non-significant. (Martin 1983) Also, yeast symbionts in cerambycid beetles did not display cellulase activity. During the 1990s, many publications identified that species of most cerambycid subfamilies could break down plant cell wall polysaccharide components including not only amorphous cellulose, but also hemicelluloses such as xylan,

β-glucan, and mannans (Martin 1991; Scrivener et al. 1997; Watanabe et al. 1998). In 2003, an enzyme of the subfamily 2 of glycoside hydrolase family 5 (GH5_2) showing endo-β-1,4-glucanase activity was identified in the yellow-spotted longicorn beetle (*Psacothea hilaris* (Pascoe)) as the first endogenous cellulase derived from a cerambycid beetle (Sugimura et al. 2003). Correlating with the increasing availability of genomic data, diverse PCWDEs have been identified in Arthropods (Acuña et al. 2012; Busch et al. 2019; Busch et al. 2017; Busch et al. 2018; Calderón-Cortés et al. 2012; Chang et al. 2012; Eyun et al. 2014; Kirsch et al. 2014; McKenna et al. 2016; McKenna et al. 2019; Pauchet and Heckel 2013; Pauchet et al. 2014; Pauchet et al. 2010; Scully et al. 2013; Shelomi et al. 2014). Interestingly, phytophaga possessed GH9, GH45, GH48, and GH28 as the common PCWDEs (Figure 6), and these enzymes have been inherited from the common ancestor of phytophaga (McKenna et al. 2019). Other PCWDEs (GH5_2, CE8, PL4, GH5_8, GH5_10, GH10, and GH11), uncommonly found in Coleoptera, also were identified in different phytophaga beetles (Figure 6) (Acuña et al. 2012; Busch et al. 2017; McKenna et al. 2016; McKenna et al. 2019; Padilla-Hurtado et al. 2012; Pauchet and Heckel 2013; Pauchet et al. 2010). These uncommon PCWDEs transferred from various microbial donors through several independent HGT events. In the next sections, I will explain in detail the three categories of PCWDEs: cellulolytic enzymes, hemicellulose enzymes, and pectolytic enzymes.

1.5.1 Cellulolytic enzymes

Cellulase are able to break down glycosidic linkage of cellulose, and are mostly glycoside hydrolases (Davies and Henrissat 1995). Cellulases are composed of three classes - endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.75 and EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21) - which are classified based on different enzymatic action and substrate specificity (Figure 7) (Watanabe and Tokuda 2010).

Endoglucanase, also known as endo-β-1,4-glucanase, cleaves β-1,4 linkages in the amorphous regions of cellulose polymers. In the case of bacterial-derived cellulase, a carbohydrate-binding domain (CBM) associates with cellulase and enhances the cellulose degradation by increasing processivity. However, insect-derived cellulases work on their substrate without a CBM. In general, endoglucanases generate mainly cellobiose and small amounts of cellotriose and glucose as the product of the enzymatic reactions (Teeri 1997).

From previous studies in Coleoptera, GH9, GH45, and GH5_2 have been identified as insect-derived endoglucanases (Busch et al. 2019; Busch et al. 2018; Davison and Blaxter 2005; Girard and Jouanin 1999a; Pauchet et al. 2014; Sugimura et al. 2003). After the first identification of endogenous GH9 cellulases in termites, enzymes of this family have been found widely in Metazoa, including in the animal phyla, Mollusca and Chordata and in plants and in some fungi (Davison and Blaxter 2005; Dehal et al. 2002; Steenbakkers et al. 2002; Suzuki et al. 2003; Urbanowicz et al. 2007). Much evidence has indicated that GH9 was present in the common ancestor of arthropods, explaining its widespread distribution amongst modern insect species (Davison and Blaxter 2005). In contrast, GH45 is predominantly identified in the microorganism (Davies et al. 1995; DeBoy et al. 2008; Sheppard et al. 1994) and has been only observed in Mollusca (Lee et al. 2004; Okmane et al. 2022; Sakamoto and Toyohara 2009), Nematoda (Kikuchi et al. 2004; Palomares-Rius et al. 2014; Wang et al. 2014; Wu et al. 2016a), and Arthropoda (Busch et al. 2019; Busch et al. 2018; McKenna et al. 2016; McKenna et al. 2019; Pauchet et al. 2014; Song et al. 2017). According to recent studies, GH45 is widely distributed in phytophagous beetles (McKenna et al. 2019) and present endoglucanse activity in some species (Busch et al. 2019; Busch et al. 2018; Ibarra et al. 2019; Valencia et al. 2013). The common ancestor of the Phytophaga is thought to have acquired GH45 horizontally from Ascomycota (Busch et al. 2019; McKenna et al. 2019). Interestingly, recent studies identified that Chrysomeloidea- and Curculionidae-derived GH45 have different substrate specificity for cellulose and hemicellulose xyloglucan (Busch et al. 2019). GH5_2 is an enzyme family predominantly found in bacteria (Aspeborg et al. 2012), but recent studies also found GH5_2 in cerambycid beetles. (McKenna et al. 2016; Pauchet et al. 2014). Cerambycid beetles have multiplied copies of GH5_2 after acquired through HGT event from a putative bacterial donor belong to Bacteriodetes (McKenna et al. 2016; McKenna et al. 2019; Pauchet et al. 2014). In addition, while bacterial-derived GH5_2 genes have been identified as mainly endoglucanases, cerambycid-derived GH5_2 has endoglucanase as well as endoxylanase activities (Pauchet et al. 2020).

Exoglucanase (also known as 1,4-β-D-glucan cellobiohydrolase or 1,4-β-D-glucan glucohydrolase) breaks down the linkage on the nonreducing or reducing ends of cellulose polymers in crystalline regions to produce cellobiose (Figure 7). Exoglucanase GH48 is commonly found in bacteria, but not usually in animals (Brunecky et al. 2017; Pauchet et al. 2014). Most bacteria possess at least one copy of

cellobiohydrolase GH48, and commonly two or three copies. However, the common ancestor of the Phytophaga acquired a gene encoding GH48 horizontally from Actinobacteria, and most extant Phytophaga harbor this enzyme (McKenna et al., 2019). The function of Phytophaga-derived GH48 has not been extensively studied yet.

β-glucosidases break the glycosidic bonds of β-D-glucosides and oligosaccharides on the non-reducing ends of cellulose polymer chains (Figure 7). These enzymes catalyze the final step in the cellulase system to produce glucose (Bhat and Bhat 1997). GH1 is a widely distributed family of β-glucosidases in bacteria, insect, and plants (Eyun et al. 2014; Ketudat Cairns and Esen 2010; Marana et al. 2001; Marques et al. 2003; Scharf et al. 2010). In the insects and microorganisms where it is present, GH1 contributes significantly to plant cell wall degradation and insect-plant interactions (Ketudat Cairns and Esen 2010; Tokuda 2019). Insect-derived GH1 originated from a common ancestor and adapted with gene duplication or loss in the insect species in evolution (He et al. 2022). GH1 has evolved with the adaptation of insects to new ecology or new feeding habits, such as the shift from plant cell walls to nectar or pollen (Johnson et al. 2018; Kunieda et al. 2006). Thus, insects may show different patterns of duplication or loss of GH1 corresponding to their lineage and ecology. GH1 genes in phytophagous beetles and aphids have independently evolved a novel role in detoxification; for example, the myrosinase genes (Beran et al. 2014; Jones et al. 2002). This newly gained function results from duplication of the ancestral insect GH1 gene; aphid myrosinase is not derived from any of the plant myrosinases. Recent studies found that phytophagous beetles possess a large number of GH1 encoding genes with high gene expression in the larval stage (Aguirre-Rojas et al. 2021; Xue et al. 2021). However, enzymatic functions of GH1 in the digestive mechanisms of phytophagous beetles are not currently well known.

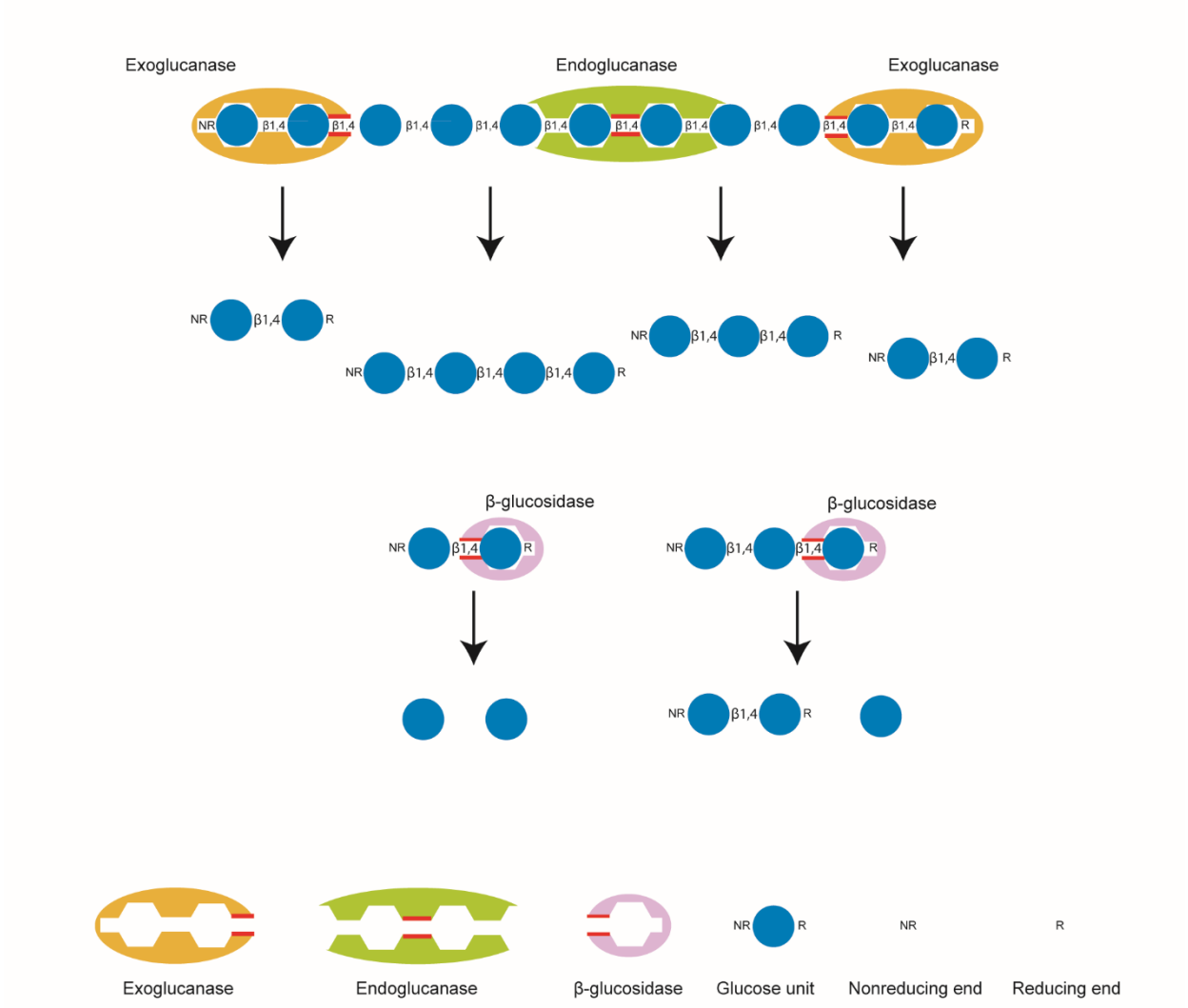

Figure 7. Schematic illustrating the enzymatic degradation of cellulose. *The three classes of cellulases - exoglucanase (EC 3.2.1.75, EC 3.2.1.91), endoglucanase (EC 3.2.1.4), and β-glucosidase (EC 3.2.1.21) – act co-operatively to break down cellulose. Exoglucanase cleaves the linkage on either non-reducing or reducing ends of cellulose polymers to produce cellobiose. Endoglucanase breaks down β-1,4 linkages in the amorphous regions of cellulose polymers. β-glucosidase completes cellulose degradation by breaking down the glycosidic bonds of β-D-glucosides and oligosaccharides on the non-reducing ends of cellulose polymers.(This figure was modified by Watanabe et al (2010))*

1.5.2 Hemicellulolytic enzymes

Hemicellulose is heterogeneous polysaccharides with a β-1,4-linked backbone of neutral sugars such as glucose, xylose, or mannose with short side chains. (Scheller and Ulvskov 2010). Hemicellulolytic enzymes (hemicellulases) break down hemicellulose polysaccharides, including xylan, xyloglucan, and gluco- and galactomannan. In contrast to cellulolytic enzymes, hemicellulolytic enzymes have a patchy distribution across the phytophagous beetles, meaning these enzymes originated from various microbial

donors via independent HGT events (McKenna et al. 2019). Hemicellulolytic enzymes allow phytophaga beetles to complete digestion by breaking down the heterogeneous structure of plant cell walls. These enzymes include xylanases, xyloglucanases, mannanases, mannosidases, and arabinosidases.

Xylanases cleave the β-1,4 linkages between xylose residues (Collins et al. 2005). They are classified into different classes based on the fold variation, enzymatic mechanism, and substrate specificity. Although all classes of xylanase are present in microorganisms, only two classes of are described here; endo-1,4-β-D-xylanases (EC 3.2.1.8) and α-arabinofuranosidases (E.C. 3.2.1.55) because my thesis focused on these enzymes in phytophaga beetles. Most previous xylanase studies focused on microorganism-derived enzymes because of producing various enzymes with multiple copy (Collins et al. 2005; Gilbert and Hazlewood 1993). GH10 and GH11 are well-studied enzymes in the microorganism: GH10 presents endo-1,4-β–xylanase (EC 3.2.1.8) endo-1,3-β-xylanases (EC 3.2.1.32), and cellobiohydrolases, whereas GH11 is known as 'true xylanase' (EC 3.2.1.8) as a monospecific catalytic active enzyme on D-xylose containing substrates (Collins et al. 2005; Drula et al. 2021). Previous studies (Padilla-Hurtado et al. 2012; Pauchet and Heckel 2013) found GH10 and GH11 in the beetles, *Hypothenemus hampei* (Curculionidae) and *Phaedon cochleariae* (Chrysomeloidea). Insect-derived GH10 catalyzes the hydrolysis of 1,4-β linkage in the backbone of xylan polymers and originated from a putative bacterial donor (McKenna et al. 2019; Padilla-Hurtado et al. 2012). GH11 in *P. cochleariae* showed enzyme activity on a xylan polymer substrate and transferred from a putative donor of proteobacteria via HGT event (McKenna et al. 2019; Pauchet and Heckel 2013). GH43 belongs to the β-Fructosidase superfamily identified with sequence similarity and presents β-xylosidases and α-L-arabinofuranosidases (Naumoff 2001). This enzyme family also has been studied in microorganisms (Cartmell et al. 2011; Suzuki et al. 2010; Valls et al. 2016; Wu et al. 2016b; Yang et al. 2015). Recent studies found that *Agrilus planipennis* (Buprestidae, Coleoptera) acquired arabinofuranosidase GH43_26 from a putative donor belonging to Actinobacteria (McKenna et al. 2019; Zhao et al. 2014). This finding is only a record for identifying the GH43 enzyme family in Coleoptera.

Endo-1,4-β-mannanase catalyzes the cleavage of β-1,4-mannosidic linkages in the mannan backbone of glucomannans and galactomannans (Malgas et al. 2015). While glucomannans consist of β-1,4-linked D-glucose and D-mannose with different ratios of these two sugars in a linear formation, galactomannan is composed of a β-1,4-linked D-mannan backbone with a single α(1,6) substitution of D-galactose.

Endomannanases have been identified in several enzyme families of glycoside hydrolases according to the CAZy database: GH5, GH9, GH26, GH44, GH113, and GH134 (Drula et al. 2021). Most endomannanase were studied in microorganisms (Dhawan and Kaur 2007; Hilge et al. 1998; Puchart et al. 2004; Talbot and Sygusch 1990), but these enzymes were also identified in different species in metazoan (King et al. 2010; Larsson et al. 2006; Shelomi et al. 2016; Song et al. 2008) and plant (Schröder et al. 2006; Yuan et al. 2007). Amongst Phytophaga, the coffee berry borer, H. hampei is reported to possess endomannanase GH5_8 (Padilla-Hurtado et al. 2012), and Callosobruchus maculatus and Gastrophysa viridula are reported to harbor mannanase GH5_10 (Busch et al. 2017). From phylogenetic analysis and sequence similarity, GH5_8 and GH5_10 have been independently horizontally transferred from bacteria - Firmicutes and Actinobacteria, respectively (Acuña et al. 2012; McKenna et al. 2019).

Endoxyloglucanases (also known as xyloglucan-specific endo-beta-1,4-glucanases, EC 3.2.1.151) cleave the xyloglucan backbone, which constitutes β-1,4-linked glucosyl backbone with a single α-1,6-D-xylosyl residue decoration (Edwards et al. 1986). Endoxyloglucanase is distributed in several enzyme families of glycoside hydrolase: GH5, GH12, GH16, GH44, and GH74 (Hasper et al. 2002; Ishida et al. 2007). Many glycoside hydrolase enzyme families act as endoxyloglucanases in bacteria and fungi (Gloster et al. 2007; Martinez-Fleites et al. 2006; Yaoi et al. 2007; Yaoi et al. 2005). However, like other hemicellulolytic enzymes, very few xyloglucanase enzymes have been identified in arthropods, specifically GH9 and GH45 (Busch et al. 2019; Shelomi et al. 2016). GH9 is a well-known ancestrally-derived cellulase, but a GH9 enzyme with xyloglucanase activity has also been reported in a stick insect (Phasmatodea) (Shelomi et al. 2016). Some GH45 enzymes in leaf beetles also have the ability to hydrolyze xyloglucan as well as cellulose. GH45 genes of phytophagous beetles made up a well-studied enzyme family and were likely inherited from the common ancestor of Phytophaga before divergence (Busch et al. 2019). Therefore, the newly-evolved substrate specificity seems to have arisen within the insect clade, likely promoted by gene duplication events.

1.5.3 Pectolytic enzymes

Pectin is composed of α-1,4-d-galacuturonic acid (GalA) residues and classified into four subclade with different structures and side chains. Depending on the cleavage site, pectolytic enzymes are classified into polygalacturonase (EC 3.2.1.15), pectinlyase (EC 4.2.2.10)/pectate lyase (EC 4.2.2.2) and pectin esterase

(EC 3.1.1.11) (Yadav et al. 2009). In my thesis, I focused on polygalacturonase because identified pectolytic enzymes in cerambycid beetles is only GH28 to date (Kirsch et al. 2014; McKenna et al. 2016; McKenna et al. 2019; Pauchet et al. 2014). In addition to cerambycid beetles, GH28 is commonly distributed in Coleoptera, Phasmatodea and Hemiptera lineages (Allen and Mertens 2008; de la Paz Celorio-Mancera et al. 2008; Eyun et al. 2014; Girard and Jouanin 1999b; Kirsch et al. 2012; Shelomi et al. 2014; Shen et al. 2003). GH28 is a polygalacturonase that cleaves α-1,4-glycosidic linkages between galacturonate residues in homogalacturonan by hydrolase mechanisms (Jayani et al. 2005; Rexová-Benková and Markovič 1976). GH28 can cleave linkages within or at the termini of the substrate (endo- or exo-cleavage). This polygalacturonase enzyme is widely distributed in most phytophagous beetles (McKenna et al. 2019). Most phytophagous beetles possess a GH28 gene which was likely transferred to the common ancestor of phytophagous beetles horizontally from a fungal donor. However, interestingly, cerambycid beetles belonging to the subfamily Lamiinae lost this copy of GH28, and acquired GH28 horizontally from a different putative fungal donor; an independent HGT event from the other phytophagous beetles (Kirsch et al. 2014).

2. Aims of the study

My work focused on studying the distribution and the evolution of plant cell wall degrading enzymes in cerambycid beetles based on transcriptome and genome data. Previous studies have already identified that two Lamiinae beetles, *A. glabripennis* and *Apriona japonica,* secrete PCWDEs into their midgut to degrade plant cell wall polysaccharides. Also, they possess families of glycoside hydrolase to break down most plant cell wall polysaccharides components. However, the family Cerambycidae consists of eight known subfamilies, and previous studies only provide information about the Lamiinae-derived PCWDEs and their evolution.

In **Manuscript 1**, I have identified PCWDE-encoding genes from the transcriptomes of 23 species representing six out of eight cerambycid subfamilies. I reconstructed a species phylogeny using orthologous genes from the transcriptome to recover evolutionary relationships between the subfamilies, which are consistent with those described in previous studies. I also curated 340 new PCWDE-encoding genes, falling into two categories: i) the broadly distributed PCWDEs originating from the common ancestor of Phytophaga (GH9, GH45, GH48, and GH28) and ii) PCWDE families that are found only in specific lineages of phytophagous beetle. With phylogenetic-based analyses, I detected putative donor lineages of horizontally acquired genes encoding PCWDEs (GH5_2, GH5_8, GH7, GH43_26, GH53, and GH10). With these findings, I provide the first comparison of PCWDEs at the subfamily level and link the distribution of PCWDEs to the evolutionary relationship of cerambycid subfamilies.

Of the cerambycid-derived PCWDEs, I focused on subfamily 2 of the glycoside hydrolase family 5 (GH5_2) because GH5_2 is not commonly identified in beetles outside of the cerambycid beetles. Also, GH5_2 found in species of the subfamily Lamiinae can degrade cellulose as well as xyloglucan and xylan and are orthologous genes. In **Manuscript 2**, I present evidence that cerambycid-derived GH5_2-encoding genes have expanded their substrate specificity through gene duplication after an HGT event. Specifically, cerambycid GH5_2 genes could break down most polysaccharide components excluding pectin. . These findings suggest that GH5_2 has played an important role in the evolution of cerambycid beetles.

Only a single genome sequence of a cerambycid beetle, *A. glabripennis* (representing the subfamily Lamiinae), has been generated to date, despite easy access to sequencing technology. A lack of genomic

information for other Lamiinae species, and other cerambycid subfamilies, has created difficulties in extending the knowledge of the evolution of PCWDEs in cerambycid beetles. Therefore, for **Manuscript 3**, I extracted high molecular weight (HMW) DNA, performed long-read sequencing, and assembled genomes for four cerambycid beetles representing three subfamilies (Cerambycinae, Lepturinae, and Lamiinae). With additional coleopteran genome data from NCBI and the project Darwin Tree of Life, I recovered the monophyletic clade of cerambycid beetles with other phytophagous beetles and Coleoptera using around a hundred orthologous genes. I also analyzed the gene structure and genomic context of newly-identified PCWDEs in cerambycid beetles (GH5_2, GH5_8, GH7, GH43_26, GH53, and GH10). I confirmed that PCWDEs originating via HGT from microbes had gained introns, and PCWDEs encoding genes were placed next to the insect genes or insect-like transposable elements. In addition, I found large conserved syntenic regions shared between different subfamilies of cerambycid beetles by comparing genome sequences. This study provides a comprehensive view of the evolution of PCWDEs in the cerambycid beetles, linked to the evolution of cerambycid beetles, by comparing draft genomes.

3. Manuscript Overview

3.1 Manuscript 1

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 D. McKenna, Yannick Pauchet

Published 11 May 2021 in Systematic Entomology

Syst Entomol (2021), 46: 784-797. doi:10.1111/syen.12488

Summary

In **Manuscript 1**, we developed a larval midgut transcriptome dataset of 23 species representing six subfamilies of cerambycid beetles. Using this transcriptome dataset, we recovered the evolutionary relationships of cerambycid beetles and identified endogenous PCWDEs. We found that cerambycid beetles possess the common phytophagous beetle PCWDEs (GH9, GH45, GH48, and GH28), as well as a number of uncommonly-identified PCWDEs (GH5_2, GH5_8, GH43_26, GH53, GH10, and GH7). Using phylogenetic analyses, we confirmed that cerambycid beetles acquired the newly-identified PCWDEs via horizontal gene transfer from different bacterial or fungal donors. This study provides the first broad view of the distribution of PCWDEs in cerambycid beetles.

Author contribution	Conceptualization: **NRS (30%)**, YP Collected specimens: PS, OD, SA Designed experiments: **NRS (70%)**, SGS, YO, BH, DDM, YP Performed experiments: **NRS (70%)**, SGS, VL, YP Data analysis: **NRS (70%)**, SGS, VL, RK, BH, DDM, YP Data visualization: **NRS (75%)**, YP Writing – original draft: **NRS (80%)**, YP Writing – review and editing: **NRS (50%)**, PS, DDM, YP

Angaben zum Eigenanteil

(gemäß der Durchführungsbestimmung zu § 8 Abs. 2 der Promotionsordnung vom 23.09.2019 der Fakultät für Biowissenschaften der FSU Jena, Anlage 2, Formular 1)

Manuscript No. 1

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

Authors: Na Ra Shin, Seunggwan Shin, Yu Okamura, Roy Kirsch, Vincent Lombard, Petr Svacha, Olivier Denux, Sylvie Augustin, Bernard Henrissat, Duane D. McKenna, Yannick Pauchet

Bibliographic information: Shin, N.R., Shin, S., Okamura, Y., Kirsch, R., Lombard, V., Svacha, P., Denux, O., Augustin, S., Henrissat, B., McKenna, D.D. and Pauchet, Y. (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. https://doi.org/10.1111/syen.12488

The candidate is

☑ First author, ☐ Co-first author, ☐ Corresponding author, ☐ Co-author.

Status: Published 11 May 2021 in Systematic Entomology

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Na Ra Shin	30%	70%	70%	50%	0%
Seunggwan Shin	0%	5%	5%	2.5%	0%
Yu Okamura	0%	5%	5%	2.5%	0%
Roy Kirsch	0%	5%	0%	2.5%	0%
Vincent Lombard	0%	0%	5%	2.5%	0%
Petr Svacha	0%	0%	0%	10%	30%
Olivier Denux	0%	0%	0%	2.5%	30%
Sylvie Augustin	0%	0%	0%	2.5%	5%
Bernard Henrissat	0%	0%	0%	2.5%	5%
Duane D. McKenna	10%	5%	0%	10%	0%
Yannick Pauchet	60%	10%	20%	12.5%	30%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

3.2 Manuscript 2

Duplication of Horizontally Acquired GH5_2 Enzymes Played a Central Role in the Evolution of Longhorned Beetles

Na Ra Shin, Daniel Doucet, Yannick Pauchet

Published June 6 2022 in Molecular Biology and Evolution

Molecular Biology and Evolution (2022) Volume 39, Issue 6, msac128, doi: 10.1093/molbev/msac128

Summary

Manuscript 2 focused on the evolution of cerambycid-specific PCWDEs, specifically the glycoside hydrolase family 5 subfamily 2 (GH5_2). Recent studies have identified multiple insect-derived GH5_2 genes in the subfamily Lamiinae of the family Cerambycidae. GH5_2 is known as a dominant cellulase in bacterial lineages, and most beetles do not harbor insect-derived GH5_2. Using a transcriptome set generated from the previous studies, we identified that most cerambycid beetles in our analysis harbor GH5_2, except for species in the subfamily Cerambycinae. We confirmed that cerambycid-derived GH5_2 has different substrate specificity than bacterial GH5_2. Using activity assays, I confirmed that cerambycid beetles could degrade most plant cell wall polysaccharides with their endogenous PCWDEs. We propose that this expansion of substrate specificities was made possible by gene duplication after horizontal gene transfer from a bacterial donor. Therefore, cerambycid-derived GH5_2 has played a significant role in the evolution of these beetles.

Author contribution

Conceptualization: **NRS (40%)** , YP
Designed experiments: **NRS (50%)**, YP
Performed experiments: **NRS (75%)**
Data analysis: **NRS (80%)**, YP
Data visualization: **NRS (100%)**
Writing – original draft: **NRS (80%)**, YP
Writing – review and editing: **NRS (65%)**, YP

Angaben zum Eigenanteil

(gemäß der Durchführungsbestimmung zu § 8 Abs. 2 der Promotionsordnung vom 23.09.2019 der Fakultät für Biowissenschaften der FSU Jena, Anlage 2, Formular 1)

Manuscript No. 2

Manuscript title: Duplication of Horizontally Acquired GH5_2 Enzymes Played a Central Role in the Evolution of Longhorned Beetles

Authors: Na Ra Shin, Daniel Doucet, Yannick Pauchet

Bibliographic information: Shin, N.R., Doucet D., Pauchet Y., Duplication of Horizontally Acquired GH5_2 Enzymes Played a Central Role in the Evolution of Longhorned Beetles, Molecular Biology and Evolution, Volume 39, Issue 6, June 2022, msac128, https://doi.org/10.1093/molbev/msac128

The candidate is

☑ First author, ☐ Co-first author, ☐ Corresponding author, ☐ Co-author.

Status: Published June 6 2022 in Molecular Biology and Evolution

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Na Ra Shin	40%	80%	75%	65%	0%
Daniel Doucet	0%	10%	5%	5%	10%
Yannick Pauchet	60%	10%	20%	30%	90%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

3.3 Manuscript 3

Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles

Na Ra Shin, Yu Okamura, Roy Kirsch and Yannick Pauchet

Submitted to Insect Molecular Biology, Status: under reviewed

Summary

Generating whole genome sequences is one of the most powerful tools available to expand our understanding of the evolution of interesting species and enzymes. In **Manuscript 3**, we generated draft genomes of four cerambycid species representing three subfamilies - Cerambycinae, Lepturinae, and Lamiinae - using Oxford Nanopore sequencing. We identified endogenous PCWDE-encoding genes using a transcriptome dataset and evolutionary relationships using phylogenetic analyses. We also used assembled genomes to investigate the evolution of uncommon PCWDEs identified in cerambycid beetles. We confirmed that newly-identified PCWDE-encoding genes in cerambycid beetles gained introns after being horizontally transferred from putative bacterial donors. Through the intron-exon structure, we stated that cerambycid-derived PCWDEs are endogenous enzymes rather than symbiont-derived and understand the evolutionary history of these enzymes in phytophaga beetles.

Author contribution	Conceptualization: **NRS (30%)** , YP Designed experiments: **NRS (60%)**, YO, RK, YP Performed experiments: **NRS (70%)** , YP Data analysis: **NRS (80%)**, YP Data visualization: **NRS (100%)** Writing – original draft: **NRS (80%)**, YP Writing – review and editing: **NRS (60%)**, YO, RK, YP

Angaben zum Eigenanteil

(gemäß der Durchführungsbestimmung zu § 8 Abs. 2 der Promotionsordnung vom 23.09.2019 der Fakultät für Biowissenschaften der FSU Jena, Anlage 2, Formular 1)

Manuscript No. 3

Manuscript title: Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles

Authors: Na Ra Shin, Yu Okamura, Roy Kirsch and Yannick Pauchet

Bibliographic information:

The candidate is

☑ First author, ☐ Co-first author, ☐ Corresponding author, ☐ Co-author.

Status: Submitted to "Insect Molecular Biology" for publication

Authors' contributions (in %) to the given categories of the publication

Author	Conceptual	Data analysis	Experimental	Writing the manuscript	Provision of material
Na Ra Shin	30%	80%	70%	60%	0%
Yu Okamura	20%	10%	15%	5%	0%
Roy Kirsch	0%	10%	15%	5%	0%
Yannick Pauchet	50%	0%	0%	30%	100%
Total:	100%	100%	100%	100%	100%

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

4. Manuscripts

4.1 Manuscript 1

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, 45075, Orléans, France
[9]Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

Published 11 May 2021 in Systematic Entomology

Systematic Entomology (2021), 46 (4), pp. 784 - 797 (2021), doi: 10.1111/syen.12488

4.3 Manuscript 2

Duplication of Horizontally Acquired GH5_2 Enzymes Played a Central Role in the Evolution of Longhorned Beetles

Na Ra Shin,[1,2] Daniel Doucet,[3] and Yannick Pauchet ,[1,2]

[1]Department of Entomology, Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany

[2]Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany

[3]Great Lakes Forestry Centre, Natural Resources Canada, Canadian Forest Service, Sault Ste. Marie, ON P6A 2E5, Canada

Published June 6 2022 in Molecular Biology and Evolution

Molecular Biology and Evolution (2022) Volume 39, Issue 6, msac128, doi: 10.1093/molbev/msac128

4.5 Manuscript 3

Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles

Na Ra Shin[1,2], Yu Okamura[1,2,3], Roy Kirsch[1,2] and Yannick Pauchet[1,2#]

[1]Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany

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

[3]Present address: Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan

Submitted to Insect Molecular Biology, Status: under reviewed

5. General Discussion

This dissertation has provided insight about the distribution and evolution of PCWDEs in subfamilies of longhorned beetles. While PCWDEs have been identified in many phytophagous beetles, including cerambycid beetles, our knowledge about the evolution and functions is mostly restricted to microorganism-derived PCWDEs. After the first endogenous cellulase in a cerambycid beetle was found (Sugimura et al. 2003), research on cerambycid-derived PCWDEs has focused on the subfamily Lamiinae (Chang et al. 2012; Kirsch et al. 2014; McKenna et al. 2016; Pauchet et al. 2014; Scully et al. 2013). Previous studies identified that species belonging to the subfamily Lamiinae harbored GH9, GH45, GH48, and GH28 (Kirsch et al. 2014; McKenna et al. 2016; Pauchet et al. 2014), which identified as common PCWDES in phytophaga beetles (McKenna et al. 2019). Interestingly, the well-known bacterial cellulase GH5_2 was also found in the subfamily Lamiinae (McKenna et al. 2016; Pauchet et al. 2014), despite the fact that there had been no previous records of this gene in phytophagous beetles (Aspeborg et al. 2012). Because most studies have focused on the subfamily Lamiinae, our understanding of the distribution of PCWDEs in cerambycid beetles has been limited. Cerambycid beetles consist of eight known subfamilies (Haddad et al. 2018) and are one of the most species-rich and abundant wood-feeding beetle famililies (Linsley 1959; Švácha and Lawrence 2014). This beetle family is an interesting model for studying adaptations for wood-feeding. Looking at more than one subfamily also allows us to determine how their complement of PCWDEs has evolved over time, rather than just seeing a snapshot in one subfamily. Thus, studying the distribution and evolution of PCWDEs in expanded subfamilies is needed and essential to our understanding of cerambycid beetles.

Looking at an expanded range of subfamilies, I investigated the distribution of cerambycid-derived PCWDEs in relation to the species phylogeny, and found that cerambycid beetles have experienced HGT events multiple times to acquire unique PCWDEs (GH5_2, GH5_8, GH43_26, GH53, GH7, and GH10). Among the cerambycid-derived PCWDEs, I proposed that a single enzyme family, GH5_2, may have played an important role in the evolution of cerambycid beetles by expanding substrate specificity through gene duplication. In order to further understand the evolution of PCWDEs in

cerambycid beetles, I searched for microsynteny in the regions next to PCWDE-encoding genes in four draft genome sequences of longhorned beetles representing three subfamilies. These studies allowed me to broaden our understanding of the distribution and evolution of PCWDEs in cerambycid beetles. Although many interesting results were found, discussions on the limitations and consequences related to my work are also necessary. I will also suggest possible future directions following the conclusions of my research.

5.1 The distribution of PCWDEs mostly follows the species phylogeny

After the first finding of putative cerambycid-derived cellulase encoding genes by protein purification and sequencing in the yellow-spotted longicorn beetle (Sugimura et al. 2003), controversy has persisted over whether this beetle is able to produce cellulase enzymes itself (Delalibera et al. 2005; Geib et al. 2009; Schloss et al. 2006). Nevertheless, cellulases (GH5, GH45) have been further identified in other species of the subfamily Lamiinae (Calderón-Cortés et al. 2010; Lee et al. 2004; Lee et al. 2005). Many beetles belonging to this subfamily have been considered severe pest species for conservation of forest ecosystems and for the forestry industry (Akbulut and Cardak 2012; Meng et al. 2015; Smith et al. 2009). Despite the family Cerambycidae being composed of eight subfamilies, research on PCWDE-encoded genes in longhorned beetles has mainly focused on the Lamiinae (Chang et al. 2012; Geib et al. 2008; Geib et al. 2010; Kirsch et al. 2014; McKenna et al. 2016; Pauchet et al. 2014; Scully et al. 2013). My project therefore provides a much broader perspective, comparing the distribution of PCWDEs in species representing six subfamilies of cerambycid beetles.

In the Cerambycidae, I found that the distribution of diverse PCWDEs followed the species phylogeny, meaning most cerambycid beetles harbored PCWDEs with other members of the Phytophaga as well as a number of less widely-distributed PCWDEs which have rarely been identified in other Phytophaga (**Manuscript 1**, Figure 2). In addition, cerambycid-derived GH5_2 has been duplicated and expanded its substrate specificity to compensate for functions that are lacking in other hemicellulases (**Manuscript 2**, Figure 6). From these results, cerambycid beetles can degrade most plant cell wall polysaccharides using endogenous PCWDE. Also, I postulated that various endogenous enzymes have evolved, complementing each other's deficient abilities for complete plant wall polysaccharide

degradation within the beetle. I hypothesized that the complementary relationships in the evolution of endogenous PCWDEs are not a trend limited to the family Cerambycidae, but can also be applied within other phytophagous beetles. Many studies have revealed and compared the function and evolution of individual PCWDEs in beetles (Busch et al. 2019; Eyun et al. 2014; Kirsch et al. 2014; Kirsch et al. 2012; Pauchet and Heckel 2013; Shelomi et al. 2014). The previous study (Busch et al. 2019; Pauchet et al. 2010) confirmed that several phytophaga beetles possess multiple copies of the GH45-encoding gene, and these multiplied enzymes can degrade cellulose, glucomannan, and xyloglucan. Busch et al confirmed that most beetles did not possess other hemicellulases that degrade mannan and xyloglucan polysaccharides, or those with hemicellulases such as GH5_8 or GH5_10 in these phytophaga beetles showed different activities from GH45. Thus, GH45 in this phytophaga beetle compensates for the deficiency of other hemicellulases with an extension of substrate specificity and increases the digestive capacity of these beetles. Thus, I hypothesize that this complementary relationship in PCWDEs is not only limited to the small number of phytophaga species or the family Cerambycidae but the general phenomenon in Phytophaga. Recent studies have shown that phytophaga beetles and Buprestoidea, a single superfamily outside phytophaga beetles, possess wide ranges of endogenous PCWDEs (McKenna et al. 2016; McKenna et al. 2019). However, research on the distribution of PCWDEs in the families Chrysomelidae and Curculionidae remains limited. In the future, investigating the broad distribution of encoded PCWDEs in other phytophagous beetles will prove crucial for understanding the relationship between PCWDE evolution and the evolution of phytophagous beetles as a whole.

5.2 Horizontally-acquired PCWDEs led to the diversification of species

Although many HGT events occur neutrally, transferred genes have been shown to adapt favorably to the recipient. In phytophagous beetles, GH28, GH45, and GH48 were transferred horizontally from different bacterial or fungal donor species (Busch et al. 2019; Kirsch et al. 2014; McKenna et al. 2019; Wybouw et al. 2016), and were inherited in Cerambycidae as common PCWDEs. I identified further horizontally-acquired genes encoding PCWDEs that seem to be unique to cerambycid beetles using phylogenetic analyses and sequence-based approaches (**Manuscript 1,** Figure 2, Figure 3). The conserved exon-intron structure of genes encoding PCWDEs, as well as syntenic regions, indicated that

these unique PCWDEs (GH5_2, GH28, GH45, and GH48) have been inherited from the most recent common ancestor of cerambycid beetles after HGT (**Manuscript 3**).

The coevolution of insects and host plants has been identified as a major driver for rapid species diversification (Ehrlich and Raven 1964). Previous studies have hypothesized that dynamic gene amplification and functional diversification of endogenous PCWDEs may have led to beetle evolution linked to host plant radiation (McKenna et al, 2019). Over the course of evolution, Cerambycid beetles have also acquired more diverse families of glycoside hydrolases through multiple HGT events (**Manuscript 1**, Figure 2) than other phytophagous beetles and other Coleoptera (McKenna et al. 2019). Symbiont-independent digestion of PCW components, made possible by these HGT events, might have helped in compensating for vulnerabilities caused by long developmental times, and promoted the diversification and success of modern cerambycid beetles (Futuyma and Moreno 1988; McKenna et al. 2019).

In conclusion, the diverse PCWDEs that Cerambycidae possess may be key to explaining how longhorned beetles have successfully evolved in a challenging environment, expanding their species diversity and abundance. Further studying the impact of diverse endogenous PCWDEs on cerambycid diversification will increase our understanding of the expansion of the Phytophaga as the second largest phytophagous lineage of insects after Lepidoptera.

5.3 Substrate specificity of endogenous PCWDEs expanded through gene duplication

Gene duplication is a significant evolutionary mechanism that can cause increasing gene dosage or expansion of substrate specificity by subfunctionalization or neofunctionalization (Innan and Kondrashov 2010). A new group of enzymes arisen from gene duplication can prefer the different substrate other than one previously favored; however, they may act on different cleavage regions that other families of enzymes that favor the same substrate (Pauchet, Ruprecht, and Pfrengle 2020). For instance, well-known xylanases GH10 and GH11 generally favor degrading relatively small arabinoxylan oligosaccharides (Collins et al. 2005; Padilla-Hurtado et al. 2012). However, in Cerambycidae, one of the xylanase GH5_2 enzymes in the subfamily Lamiinae exhibits different enzymatic activity, degrading xylohexaose or longer xylan oligomers (Pauchet et al. 2020). Therefore,

as the result of gene duplication after HGT, cerambycid-derived GH5_2 evolved the ability to degrade most polysaccharide components of plant cell wall, rather than only breaking down cellulose (**Manuscript 2**, Figure 1).

GH5_2 is not the only case of expanded substrate specificity due to gene duplication. From considering the large number of genes encoded in the transcriptome and genome relative to other beetle species, I found that cerambycid-derived GH45 encoding genes are also the result of gene duplication (**Manuscript 1**, Figure 2). In Phytophaga, the increased number of GH45 encoding genes was linked to the diversification of substrate specificity to include xyloglucan, cellulose, and glucomannan (Busch, Danchin, and Pauchet 2019). Interestingly, four species in the subfamily Prioninae and *Molorchus* minor, a species of the subfamily Cerambycinae, do not possess GH5_2 or any other hemicellulolytic enzymes in cerambycid beetles (**Manuscript 1**, Figure 2), and they possess more than five and up to eight genes encoding GH45 (**Manuscript 2**, Figure 6). This result presents the high possibility that GH45 in these five species has evolved to complement the absence of hemicellulase in cerambycid beetles like GH45 in several phytophaga beetles (Busch et al. 2019).

In my research, I investigated four species of the subfamily Prioninae, all collected from Europe (**Manuscript 1**, Table 1). These four species only harbor GH9, GH45, GH48, and ancestral type GH28, inherited from the most ancestor of the Phytophaga. Interestingly, all four species present high copy number of GH45-encoding genes, ranging from five to eight copies. *M. minor* also shows a similar pattern, having an increased copy number of GH45-encoding genes without other known hemicellulases. These five species might have adapted to the lack of any hemicellulases by expanding substrate specificity of GH45 by gene duplication and neofunctionalization, allowing for complete digestion of plant cell wall polysaccharides. To clarify cerambycid-derived GH45 evolution will require further research to identify enzyme activity; moreover, studying the activities of other PCWDEs in longhorned beetles may reveal further diversification of enzymatic functions. Although cerambycid beetles only harbor a single GH9-encoding gene, stick insects possess many genes encoding GH9, an enzyme able to degrade most PCW components except for mannose and poly- and oligosaccharides (Shelomi et al.

2016). Therefore, further studies on the remaining PCWDEs may provide new insight into the evolution of cerambycid-derived enzymes.

5.4 Cerambycidae is comprised of eight subfamilies, with two being relatively understudied

To identify genes encoding glycoside hydrolase families in cerambycid beetles, I analyzed 23 species from six out of eight subfamilies (**Manuscript 1**). The remaining two subfamilies, Dorcasominae and Parandrinae, could unfortunately not be included my analysis. McKenna et al, found that a species within Parandrinae harbored GH9, GH28, GH45, and GH48 as the common PCWDEs in the Phytophaga (McKenna et al. 2019). In the phylogenetic analysis, Parandrinae was revealed to be a sister subfamily to Prioninae (Haddad et al. 2018), and studied species belong to these two subfamilies neither harbor any hemicellulolytic enzymes nor PCWDEs acquired from recent HGT events (**Manuscript 1**, Figure 2) (McKenna et al. 2019). This distribution of these digestive enzymes still support a correlation between the distribution of PCWDEs and the evolutionary relationships of cerambycid beetles. The remaining subfamily Dorcasominae does not have any record of PCWDEs to date. Dorcasominae was shown to be a sister subfamily to Cerambycinae (Haddad et al. 2018). A better understanding of the connections between digestive enzyme distribution and evolutionary relationships within cerambycid beetles requires further research, namely collecting species from the remaining two subfamilies and studying their endogenous PCWDEs. Dorcasominae, with over 300 species, is mainly distributed in Madagascar (Švácha and Lawrence 2014). Parandrinae, with fewer species (~113), is mainly found in central and south America, although it has representatives worldwide. Both subfamilies have been considered of southern origin like Cerambycinae and Lepturinae (Švácha and Lawrence 2014). I investigated 23 species mostly collected inside Europe except for a single species, *Cacosceles newmannii,* from South Africa (Manuscript 1, Table 1). Therefore, expanding the collecting regions from not only Europe to southern regions, especially Africa, may help to complete the set of species for eight known subfamilies in cerambycid beetles. Nevertheless, *Neandra brunnea*, a species belonging to the subfamily Parandridinae, has been recorded in Dresden (Saxony, Germany) in 2014. Therefore, further work to collect beetles in Europe could be still considered as the way to complete the remaining subfamilies.

5.5 Chromosome-level genomes are essential to understand their evolution.

In genomic studies of arthropods, Diptera and Lepidoptera are over-represented in terms of insect groups with assembled genomes (Hotaling et al. 2021). Although Coleoptera has also been included in genome projects for arthropods, such as i5k and the Darwin Tree of Life Project (Blaxter et al. 2022; Poelchau et al. 2015), the draft genome sequence of *A. glabripennis* was the only available sequence of a cerambycid beetle publicly available (McKenna et al. 2016). To compare the evolution of PCWDEs in cerambycid beetles, I needed genome data set at least representing three subfamilies (Manuscript 3). I employed Nanopore sequence technology to generate the long-read sequence. Oxford Nanopore MinION sequencing is prone to high-error rates between 10 to 15 percent during generating long-read compared to other sequence technology such as Illumina (Lu et al. 2016). To reduce the error rates in genome assembly, compensating long-read from Oxford Nanopore technology and short-read from Illumina sequence are recommended to fix results from misassembly and sequence errors (Chen et al. 2021). Also, Hi-C can correct contig misalignments and scaffold contigs into chromosomes on the long-read genome with Hi-C library sequenced on Illumina sequence (Yuan et al. 2017).

However, our sample size is limited due to the usage of rest body tissue from extracting RNA seq for transcriptome analysis. I could not employ Hi-C technology or Illumina sequence because I did not have enough species materials for Illumina sequencing. To decrease the error rate from nanopore sequencing, I excluded a short read for preparing the gDNA library for sequencing and facilitated the specific assembler, Flye, to improve the quality of draft genome assembly. Unlike another assembler, Flye generates high contiguity of the genome by concatenating all error-prone to increase the accuracy of the sequence (Kolmogorov et al. 2019). This is likely why the four draft genomes I generated using long-read sequences (**Manuscript 3**) did not reach chromosome-level quality, despite high scores from the BUSCO assessment. High heterozygosity levels are also one reason to achieve high-quality assembly (Li et al. 2019). In other insect species, Hi-C has been shown to dramatically improve the ability to generate chromosome-length scaffolds (Dudchenko et al. 2017; Ellison and Cao 2019). Draft genomes reveal basic information on genetic properties, but cannot provide a full understanding of the evolution of a species; a chromosome-level genome is crucial for the study of genome structure and comparison.

Recently, research within the Darwin Tree of Life project generated a high-quality, chromosome-level genome for *Rutpela maculata*, belonging to the cerambycid subfamily Lepturinae (INSDC ID: PRJEB51449). I compared the draft genomes of two species representing the subfamilies Lamiinae and Cerambycinae to the chromosome-level genome of *R. maculata* (**Manuscript 3**, Figure 6). However, comparing the conserved regions including PCWDEs proved complicated due to dynamic genome rearrangement and the insufficient lengths of contigs. In the future, chromosome-level whole genome sequences for species within the Cerambycidae will be necessary to expand our understanding of the evolution of this lineage of beetles, especially in comparing between the subfamilies.

5.6 Most studies on PCWDE structure and catalytic mechanisms focus on microorganism-encoded enzymes

GH5_2 has been found only in cerambycid beetles (McKenna et al. 2016; Pauchet et al. 2014), meaning most beetle species do not possess this endogenous PCWDE (McKenna et al. 2019). Before finding this enzyme in longhorned beetles, GH5_2 was predominantly known as a bacterial enzyme, so most studies on protein structure and catalytic sites have focused on bacterial GH5_2 (Aspeborg et al. 2012; Zhu et al. 2016). I employed a protein data bank (PDB) model of bacterial cellulase as a reference to attempt the homology modeling of cerambycid GH5_2. Interestingly, I found that cerambycid GH5_2 enzyme-encoding genes share the commonly recognized active sites of bacterial cellulase GH5_2, regardless of substrate specificity (**Manuscript 2**, Figure 5). Conserved active sites are significant markers indicating the essential information for enzymes' property and substrate specificity. Therefore, the replacement of active site residues can cause property changes or weakened enzyme ability (Lee et al. 2011; Liu et al. 2010; Zhu et al. 2016). Diversified substrate specificity in cerambycid-derived GH5_2 does not show any association with conserved amino acid residues on active sites; indeed, transglycosylase GH5_2 has the same amino acid residues as the bacterial cellulase GH5_2 in its active site, despite their very different substrate specificities (**Manuscript 2**, Figure 5, and Figure S9). Therefore, our view on the evolution of cerambycid-derived GH5_2 may be limited by relying solely on existing enzyme research from a bacteria. In previously known studies, I can find examples like GH5_2 in cerambycid beetles. In the case of GH45 from phytophaga beetle, this enzyme retained the same conformation as a

conserved catalytic domain known from fungal cellulase studies while exhibiting an extended substrate for plant cell wall polysaccharides (Busch et al. 2019). Bacterial CBM35 also exhibited different substrate specificities and biological functions despite highly conserved amino acid residues at the ligand binding site (Montanier et al. 2009). However, many studies still stated that conserved catalytic residues related to the substrate specificity (Bartlett et al. 2002; Ribeiro et al. 2020; Salmon et al. 2016). In order to understand the relationship between the catalytic regions and divergent substrate specificity of cerambycid-derived GH5_2, it is necessary first to confirm the more accurate protein structure using crystallization and to identify the specific catalytic regions of this enzyme through site mutagenesis.

5.7 Studying the effect of enzyme function *in vivo* with genetic approaches

To clarify the role of PCWDEs *in vivo*, gene silencing (RNA interference) or knockout (CRISPR/Cas9) are effective methods in a target species. In early 1990, Seiboth et al, employed gene disruption by homologous recombination to study the effect of cellulolytic enzymes. In the study about the filamentous fungus *Trichoderma reesei* (Hypocreaceae), which secretes large amounts of cellulolytic enzymes, the absence of cellulase inhibited or prevented fungal growth (Seiboth et al. 1992). Compared to studies of PCWDEs in fungi, only a few reports have studied the effects of PCWDEs in insects *in vivo*. *Phaedon cochleariae* (Curculionoidea, Coleoptera) have been studied using RNAi and CRISPR/CAS9. (Kirsch et al. 2019; Kirsch 2022). In a study of this beetle, downregulating inactive GH28 caused high expression of active endo PG GH28, whereas silencing genes encoding active endoPG GH28 did not impact the expression of other GH28 genes or of other endogenous PCWDE-encoding genes. Also, silencing endoPG GH28 encoding genes had no apparent effect on insect growth (Kirsch et al. 2019). However, complete knockout of endoPG GH28 in the same species using CRISPR/Cas9 caused adverse effects on insect growth and increased mortality (Kirsch 2022). Cerambycid beetles possess diverse PCWDEs, which have likely evolved to compensate for the inability of other digestive enzymes to completely degrade PCW polysaccharides (**Manuscript 1**, Figure 2, and **Manuscript 2**, Figure 6). However, how various plant wall degrading enzymes affect cerambycid beetles is not known. Thus, gene silencing or knockout may be the first step to shedding light on the relationships and evolution of endogenous digestive enzymes. Because cerambycid beetles are invasive pest species, most studies have focused on adjusting mortality of beetles using RNA

interference (RNAi). In addition, since most studies have aimed to apply the technique to the field, dietary methods for inducing gene silencing have been the primary method studied (Dhandapani et al. 2020b; Willow and Veromann 2021).

The Cerambycidae is not an easy family of beetle to rearing due to long developmental times for the larval stages (Linsley 1959). After hatching from eggs, longhorned beetle larvae bore into the host plant, remaining inside it for several years before they develop into adults (Švácha and Lawrence 2014; Wang 2017). Prioninae and Parandrinae prefer decaying wood for their larval developmental stage, whereas Cerambycinae and Lamiinae accept living or dead trees and even wood products (Cocquempot and Lindelöw 2010; Raje et al. 2016). Therefore, due to the long developmental periods of cerambycid beetles, knockdown experiments have only been carried out from 3 days to 8 months during the early stage of larval development (Dhandapani et al. 2020a; Dhandapani et al. 2020b). Cerambycid beetles are also challenging to reproduce in the lab because these beetles prefer to breed at night and are sensitive to environmental changes (Keena 2017; Linsley 1959). Dhandapani et al. succeeded in breeding and rearing cerambycid beetles with freshly cut host trees in the lab (Dhandapani et al. 2020b). If beetles belonging to other subfamilies rather than Lamiinae could be reared and bred, I speculate that the effect of enzyme loss can be clearly understood using RNAi and CRISPR/Cas9. I characterized the enzyme activity of GH5_2 on different substrates *in vitro* (Manuscript 2, Figure S4-S8), but the roles of this enzyme family in cerambycid beetles *in vivo* was not studied yet. Therefore, *in vivo* studies on knocking out GH5_2 via CRISPR/Cas9 in cerambycid beetles and investigating their growth or any effects on their mortality may lead to understanding the function and distribution of PCWDEs in the evolutionary history of these beetles. To better understand the role of PCWDEs in cerambycid beetles, rearing and breeding issues and inefficiencies must be resolved in the future.

5.8 Insect-derived PCWDEs may contribute to sustainable energy applications

In biotechnology research, the ability of certain enzymes to break down plant polymers has been used to develop renewable substitutes for fossil fuels, mainly by using fungal and bacterial cellulases to degrade lignocellulose into liquid fuels such as ethanol (Wilson 2009). Microorganism-derived xylanases also play a significant role in many biotechnological applications (Subramaniyan and Prema

2002). For instance, xylanases break down hemicellulose without degrading cellulose in the pulping process of paper production, maintaining cellulose structure and therefore paper quality (Kaur et al. 2016; Shoham et al. 1992). Alkaliphilic and acidophilic xylanases are also essential for bleaching or detergent applications (Beg et al. 2001; Viikari et al. 1994) However, to increase the yield of degrading enzymes, preparation steps are required to activate cellulase or generate cellulase-free xylanase with prior heat treatment (Chang et al. 2016; Shrinivas et al. 2010). Finding appropriate enzymes that can generate high-yield products with low cost is still a significant challenge for the biotechnology industry.

In the biotechnology industry, cellulase, xylanase, pectinase, and enzymes that can degrade lignin have been considered in applications ranging from generating biofuel, recycling paper, to food processing (Prajapati et al. 2018). To date, degrading enzymes isolated from bacteria or fungi were commonly chosen for biotechnological applications because they are commonly found in nature. The habitats of microorganisms vary widely, and isolated microorganism-derived enzymes work on different environmental conditions such as temperature and pH for activity (Thapa et al. 2020). However, the growing demand for more efficient degrading enzymes has led to research on new PCWDEs derived from different organisms, including animals.

Individual degrading enzymes have been employed as essential tools in biotechnology, but in nature these enzymes exhibit efficient activity while in synergistic relationships (Bayer et al. 2004). Research on multi-enzyme synergism can even advance biofuel production (Bhattacharya et al. 2015). Cerambycid species possess diverse endogenous PCWDEs, which have evolved in complementary relationships, and studying this synergism could result in new approaches to reducing chemical waste. Cerambycid-derived PCWDEs work in the same gut environment between pH 6 to 7.5 (Holtof et al., 2019), meaning these enzymes have already adapted to the same environment to compensate for the absent function of other enzymes. Studying the synergistic relationships between GH5_2 and GH10, which target different cleavage sites of, may contribute to developing new biotechnological applications. Understanding enzyme properties and mechanisms when in synergistic relationships may therefore bring new insight into biotechnological applications.

One cerambycid PCWDE with great potential for biosynthetic applications is transglycosylase.. Historically, many methods were developed to generate transglycosylase from glycoside hydrolase enzymes (Bissaro et al. 2015). Since retaining glycoside hydrolases follow a double displacement mechanism, synthetic activity has been induced by shifting the balance either towards glycosidic bond formation or using an activated glycosyl donor (Planas and Faijes, 2002). However, artificially produced transglycosylases often displayed low levels of transglycosylation activity, because the main products were still generated from hydrolase activity rather than synthesis activity. Ceramycid transglycosylase enzyme prefers synthesizing glycosidic bonds to hydrolysing them without any pretreatment (**Manuscript 2**, Figure 4). Studying this cerambycid-derived transglycosylase further may reveal the mechanism of transglycosylase derived from hydrolase activity and allow its application to other hydrolase enzymes.

5.9 Conclusion

Cerambycid beetles possess endogenous PCWDEs so that they can degrade polysaccharide components and acquire nutrients in challenging environments. The research within this dissertation has furthered our understanding of the evolution of PCWDEs in cerambycid beetles and helped elucidate a possible role of endogenous PCWDEs in promoting species diversification within Cerambycidae. While cerambycid beetles harbor diverse PCWDEs, I have only functionally analyzed one enzyme family, GH5_2. Studying the remaining endogenous digestive enzymes in cerambycid beetles would complement this work. Characterizing GH45 allows us to understand how one common enzyme has evolved to expand its ability to degrade plant cell walls without additional enzymes acquired from HGT events. Studying the role of diverse PCWDEs in cerambycid beetles *in vivo* also provides evidence of the successful evolution of cerambycid beetles without symbiont association. In addition, our understanding of the evolution of PCWDEs in the complementary relationship will expand our knowledge of how the family Cerambycidae, specifically, and other phytophagous beetles generally, have adapted to feed on plant tissues.

6. Summary

The Cerambycidae, also known as longhorned beetles, are a family of plant-feeding beetles (Phytophaga) that specifically feed on wood (xylophagy). Woody tissues, which are mainly composed of plant cell wall polysaccharides like cellulose, hemicellulose, pectin, and lignin, lack essential nutrients for insect development, and the structure of long-chain polysaccharides makes it difficult for beetles to degrade the plant cell wall within their digestive systems. However, Cerambycidae have evolved in this challenging environment by successfully acquiring nutrients from woody tissues. In order to break down polysaccharides, longhorned beetles depend on plant cell wall degrading enzymes (PCWDEs). Before the discovery of insect-derived cellulases, it was assumed that symbionts, like bacteria and yeasts, provided the beetles with these digestive enzymes. More advanced technology for conducting genomic studies has since allowed the identification of insect-derived PCWDEs in many insect species. Within the Cerambycidae family, consisting of eight subfamilies, research on PCWDEs has focused on only one subfamily (Lamiinae). Therefore, my project focused on studying PCWDEs in several Cerambycidae subfamilies, investigating how these enzymes are distributed and how they have contributed to the evolution of this insect family.

I identified PCWDEs secreted from the midguts of larvae, and aimed to analyze their evolution within the Cerambycidae family. For this purpose, I generated a larval midgut transcriptome dataset of 23 cerambycid species representing six subfamilies: Cerambycinae, Prioninae, Lepturinae, Necydalinae, Spondylidinae, and Lamiinae. I inferred a species phylogeny of cerambycid beetles using orthologous genes detected from transcriptomes, which confirmed the previously described evolutionary relationships of the family Cerambycidae. Through analyzing the cerambycid transcriptomes, I could then describe the number and distribution of cerambycid-derived PCWDEs. I identified 340 PCWDE-encoding genes, including enzymes associated with phytophagy in other beetle species, as well as enzymes unique to cerambycid beetles. Activity assays also showed that longhorned beetles could break down most plant cell wall polysaccharides with their endogenous PCWDEs. Using phylogenetic analyses, I investigated the evolution of cerambycid PCWDEs and putative donor lineages in the context of recent horizontal gene transfer (HGT) events. Based on these analyses, I speculated that the

distribution of PCWDEs follows species phylogeny, not relationships to host species or ecological habitats, meaning larval host plant did not related to the evolutionary history of PCWDEs in the cerambycid beetles.

The PCWDE GH5_2 is commonly found in bacteria, but not in beetles. However, previous studies on the subfamily Lamiinae revealed that they also possess GH5_2 enzymes, unlike other lineages of beetles. GH5_2 enzymes encoded by genes identified in two species of Lamiinae also have been shown to break down cellulose as well as the hemicelluloses xyloglucan and xylan. These genes were shown to be orthologous in the phylogenetic analysis. I have found that this enzyme is distributed throughout cerambycid beetles except most species of the subfamily Cerambycinae and all species in the subfamily Prioninae in 23 collected species. Phylogenetic analysis indicated that cerambycid beetles have acquired GH5_2 via HGT from a bacterial donor.

Using enzyme activity assays, I confirmed that cerambycid-derived GH5_2 enzymes break down all tested PCW polysaccharides except pectin. The evolution of this substrate specificity expansion was facilitated by gene duplication. To discover the function of the ancestral GH5_2 gene, I reconstructed the sequences of the most likely common ancestral GH5_2 genes using phylogenetic reconstruction methods. The most likely reconstructed ancestral GH5_2 showed promiscuous activities to cleave β-1,4 glycosidic bonds in cellulose, glucomannan, and xylan. This indicates that gene duplication after HGT allowed neofunctionalization and subfunctionalization of ancestral GH5_2 encoding genes, which expanded their substrate specificity to allow degradation of most PCW polysaccharides.

Although current access to genome sequencing is facilitated by cheaper prices and advanced technologies, comparatively few draft genomes have been published for cerambycid beetles, and no model species with chromosome-level genomes yet exist. I generated long-read sequences using small amounts of insect body tissues to understand the evolution of PCWDEs at the species level; I selected four cerambycid species representing three subfamilies: Cerambycinae, Lepturinae, and Lamiinae. I then generated a set of high-quality draft genomes for the four cerambycid species, and inferred the phylogenetic relationships between longhorned beetles and other beetles using orthologous genes.

Although none of the four draft genomes met the quality standard for chromosome level, synteny comparison between them allowed me to analyze conserved PCWDE-related regions between species.

Although this thesis has contributed significantly to our understanding of the evolution and the distribution of PCWDEs in cerambycid beetles, further studies are needed to complete understanding. I could not include the two remaining subfamilies, Dorcasominae and Parandrinae because of low number of extant species and limited rearing area in the world. Therefore, investigating endogenous PCWDEs in species of these two subfamilies may bring the help to understand completely the distribution and properties of PCWDEs in Cerambycidae. Further experiments, such as protein crystallization and mutagenesis, also will be required to understand the mode of action of PCWDEs' active sites residues and digestive mechanisms in beetles.

I provided the first comprehensive study of genomic data from the family Cerambycidae, using 23 species of beetles representing six subfamilies. I have shown that most cerambycid beetles are not fully dependent on microorganisms or symbionts to digest plant matter, but possess their own set of digestive enzymes. Enzyme-encoding genes have been horizontally acquired from various microbial donors, and have evolved by gene duplications after these HGT events occurred. These findings are an excellent starting point for studying the complementary function and evolution of endogenous PCWDEs in cerambycid and other phytophagous beetles.

7. Zusammenfassung

Die Cerambycidae, auch Bockkäfer genannt, sind eine Familie pflanzenfressender Käfer (Phytophaga), die sich speziell von Holz ernähren (Xylophagie). Dem Holzgewebe, das hauptsächlich aus pflanzlichen Zellwandpolysacchariden wie Zellulose, Hemizellulose, Pektin und Lignin besteht, fehlen die für die Entwicklung der Insekten essenziellen Nährstoffe und die Struktur der langkettigen Polysaccharide macht es den Käfern zusätzlich schwer, Zellwandbestandteile in ihrem Verdauungssystem abzubauen. Cerambycidae haben sich jedoch erfolgreich in diesem schwierigen Habitat entwickelt, da sie effizient Nährstoffe aus holzigen Geweben gewinnen können. Um Polysaccharide abzubauen, sind Bockkäfer auf Pflanzenzellwand-abbauende Enzyme (PCWDEs) angewiesen. Vor der Entdeckung der insekteneigenen Zellulasen nahm man an, dass symbiotische Bakterien und Hefen die Käfer mit diesen Verdauungsenzymen versorgen. Dank technologischer Fortschritte in der Genomanalyse konnten inzwischen weitere endogene PCWDEs in vielen Insektenarten nachgewiesen werden. Innerhalb der Familie der Cerambycidae, die aus insgesamt acht Unterfamilien besteht, konzentrierte sich die Forschung an PCWDEs bisher nur auf eine Unterfamilie (Lamiinae). Meine Doktorarbeit beschäftigt sich daher gezielt mit der Verbreitung von PCWDEs in weiteren Unterfamilien der Cerambycidae und wie diese Enzyme zur Evolution dieser Insektenfamilie beigetragen haben könnten.

Zunächst habe ich PCWDEs aus Larvendarmmaterial isoliert und ihre Evolution innerhalb der Familie der Cerambycidae untersucht. Hierfür habe ich einen Transkriptomdatensatz von 23 Cerambycidae-Arten aus sechs Unterfamilien generiert: Cerambycinae, Prioninae, Lepturinae, Necydalinae, Spondylidinae und Lamiinae. Anhand von orthologen Genen, die ich in den Transkriptomen nachweisen konnte, habe ich eine Artenphylogenie der Cerambycidae abgeleitet. Diese Phylogenie bestätigte die zuvor beschriebenen evolutionären Beziehungen in der Cerambycidae-Familie. Durch die Transkriptomanalyse konnte ich dann die Anzahl und Verbreitung endogener Cerambycidae-PCWDEs beschreiben. Ich habe 340 PCWDE-kodierende Gene identifiziert, darunter Enzyme, die bei anderen Käferarten mit Phytophagie in Verbindung gebracht werden, sowie Enzyme, die es ausschließlich in Cerambycidae gibt. Aktivitätsanalysen zeigten außerdem, dass Bockkäfer die meisten Pflanzenzellwandpolysaccharide mit ihren endogenen PCWDEs abbauen können. Mit Hilfe von

weiteren phylogenetischen Analysen habe ich die Entwicklung der Cerambycidae-PCWDEs und der mutmaßlichen Donorlinien im Hinblick auf den jüngsten horizontalen Gentransfers (HGT) untersucht. Die Analyse deutet daraufhin, dass die Verbreitung der PCWDEs der Phylogenie der Arten folgt und nicht in Beziehungen zu den Wirtspflanzenarten oder den ökologischen Lebensräumen steht. Dies bedeutet auch, dass die jeweilige Wirtspflanze während der Larvenentwicklung wahrscheinlich keinen Einfluss auf die Evolutionsgeschichte der Cerambycidae-PCWDEs hat.

PCWDE GH5_2 kommt häufig in Bakterienarten vor, aber eigentlich nicht in Käfern. Frühere Studien über die Unterfamilie Lamiinae ergaben jedoch, dass sie im Gegensatz zu anderen Käferlinien ebenfalls GH5_2-Enzyme besitzen. GH5_2-Enzyme zweier Lamiinae-Arten bauen nachweislich auch Zellulose sowie die Hemizellulosen Xyloglucan und Xylan ab. Diese Gene erwiesen sich in der phylogenetischen Analyse als ortholog. Ich habe festgestellt, dass diese Enzyme bei allen Cerambycidae außer den meisten Arten der Unterfamilie Cerambycinae und der Unterfamilie Prioninae in insgesamt 23 gesammelten Arten vorkommen. Phylogenetische Analysen ergaben außerdem, dass Cerambycidae-Käfer GH5_2 Enzyme über HGT von einem bakteriellen Donor erworben haben.

Mit Hilfe von Enzymaktivitätstests konnte ich bestätigen, dass die von Bockkäfern stammenden GH5_2-Enzyme alle getesteten PCW-Polysaccharide außer Pektin abbauen. Die Erweiterung der Substratspezifität dieser Enzyme wurde durch Genduplikation ermöglicht. Um die ursprüngliche Funktion des GH5_2-Gens zu ermitteln, habe ich die Sequenzen der wahrscheinlichsten gemeinsamen GH5_2-Gene mit Hilfe phylogenetischer Methoden rekonstruiert. Das rekonstruierte GH5_2-Gen zeigte vielseitige Aktivitäten zur Spaltung von β-1,4-glykosidischen Bindungen in Zellulose, Glucomannan und Xylan. Dies deutet darauf hin, dass die Genduplikation nach HGT eine Neofunktionalisierung und Subfunktionalisierung der für GH5_2 kodierenden Gene ermöglichte, was wiederum die Erweiterung des Substratspektrums auf die meisten PCW-Polysaccharide zur Folge hatte.

Obwohl die Anzahl der Genomsequenzierungen durch reduzierte Kosten und fortschrittliche Technologien deutlich gestiegen ist, sind bislang nur vergleichsweise wenige Genome für Cerambycidae-Käfer veröffentlich und keine der Arten dient der Forschung bisher als Modellorganismus. Ich habe Long-Read-Sequenzen aus kleinen Mengen von Insektenkörpern generiert, um die Evolution von PCWDEs auf Artniveau zu verstehen; ich habe vier Cerambycidae-Arten

ausgewählt, die drei Unterfamilien repräsentieren: Cerambycinae, Lepturinae und Lamiinae. Anschließend habe ich einen Datensatz hochwertiger Genom-Entwürfe für die vier Cerambycidae-Arten erstellt. Anhand orthologer Gene ließ sich auf die phylogenetischen Beziehungen zwischen Bockkäfern und anderen Käfern schließen. Obwohl keiner der vier Genom-Entwürfe den Qualitätsstandard für die Chromosomenebene erfüllte, ermöglichte mir der Syntenie-Vergleich, konservierte PCWDE-bezogene Regionen zwischen den Arten zu analysieren.

Obwohl diese Arbeit einen wichtigen Beitrag zu unserem Verständnis der Evolution und der Verbreitung von PCWDEs bei Cerambycidae-Käfern geleistet hat, sind weitere Studien erforderlich. Die beiden verbleibenden Unterfamilien, Dorcasominae und Parandrinae, konnte ich aufgrund der geringen Artenzahl und des sehr begrenzten weltweiten Vorkommens nicht in die Analysen miteinbeziehen. Die Untersuchung endogener PCWDEs in diesen beiden Unterfamilien könnte dazu beitragen, die Verbreitung und Eigenschaften von PCWDEs in Cerambycidae vollständig zu verstehen. Um die Wirkungsweise und Funktion der PCWDEs und die Verdauungsmechanismen in Käfern besser zu verstehen, könnte z. B. mittels Proteinkristallisation oder Mutagenese das aktive Zentrum der Enzyme in zukünftige Experimenten untersucht werden.

Im Rahmen dieser Doktorarbeit habe ich die erste umfassende Studie von Genomdaten aus der Familie der Cerambycidae vorgelegt, die 23 Käferarten aus sechs Unterfamilien umfasst. Ich habe gezeigt, dass die meisten Cerambycidae-Käfer bei der Verdauung pflanzlicher Stoffe nicht vollständig von Mikroorganismen oder Symbionten abhängig sind, sondern eigene, endogene Verdauungsenzyme besitzen. Enzymkodierende Gene wurden horizontal von verschiedenen mikrobiellen Donatoren erworben und haben sich durch Genduplikation weiterentwickelt. Diese Ergebnisse sind ein guter Ausgangspunkt für die weitere Untersuchung der komplementären Funktionsweise und der Evolution endogener PCWDEs bei Cerambycidae und anderen phytophagen Käfern.

8. References

Abbott S, Fairbanks DJ. 2016. Experiments on plant hybrids by gregor mendel. Genetics. 204(2):407-422.

Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH, Egan AN, Doyle JJ. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proceedings of the national academy of sciences. 109(11):4197-4202.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF. 2000. The genome sequence of drosophila melanogaster. Science. 287(5461):2185-2195.

Aguirre-Rojas LM, Scully ED, Trick HN, Zhu KY, Smith CM. 2021. Comparative analyses of transcriptional responses of dectes texanus leconte (coleoptera: Cerambycidae) larvae fed on three different host plants and artificial diet. Scientific reports. 11(1):1-19.

Akbulut Y, Cardak CS. 2012. Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011. Computers & Education. 58(2):835-842.

Allen ML, Mertens JA. 2008. Molecular cloning and expression of three polygalacturonase cdnas from the tarnished plant bug, lygus lineolaris. Journal of Insect Science. 8(1).

Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (gh5). BMC evolutionary biology. 12(1):1-16.

Bacic A, Harris PJ, Stone BA. 1988. Structure and function of plant cell walls. The biochemistry of plants. 14:297-371.

Barnes WJ, Anderson CT. 2018. Release, recycle, rebuild: Cell-wall remodeling, autodegradation, and sugar salvage for new wall biosynthesis during plant development. Molecular Plant. 11(1):31-46.

Bartlett GJ, Porter CT, Borkakoti N, Thornton JM. 2002. Analysis of catalytic residues in enzyme active sites. Journal of molecular biology. 324(1):105-121.

Bayer EA, Belaich J-P, Shoham Y, Lamed R. 2004. The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol. 58:521-554.

Beg Q, Kapoor M, Mahajan L, Hoondal G. 2001. Microbial xylanases and their industrial applications: A review. Applied microbiology and biotechnology. 56(3):326-338.

Beran F, Pauchet Y, Kunert G, Reichelt M, Wielsch N, Vogel H, Reinecke A, Svatoš A, Mewis I, Schmid D. 2014. Phyllotreta striolata flea beetles use host plant defense compounds to create their own glucosinolate-myrosinase system. Proceedings of the National Academy of Sciences. 111(20):7349-7354.

Berglund J, Angles d'Ortoli T, Vilaplana F, Widmalm G, Bergenstråhle-Wohlert M, Lawoko M, Henriksson G, Lindström M, Wohlert J. 2016. A molecular dynamics study of the effect of glycosidic linkage type in the hemicellulose backbone on the molecular chain flexibility. The Plant Journal. 88(1):56-70.

Bhat M, Bhat S. 1997. Cellulose degrading enzymes and their potential industrial applications. Biotechnology advances. 15(3-4):583-620.

Bhattacharya AS, Bhattacharya A, Pletschke BI. 2015. Synergism of fungal and bacterial cellulases and hemicellulases: A novel perspective for enhanced bio-ethanol production. Biotechnology letters. 37(6):1117-1129.

Biely P. 2012. Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnology Advances. 30(6):1575-1588.

Biémont C. 2010. A brief history of the status of transposable elements: From junk DNA to major players in evolution. Genetics. 186(4):1085-1093.

Bissaro B, Monsan P, Fauré R, O'Donohue MJ. 2015. Glycosynthesis in a waterworld: New insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. Biochemical Journal. 467(1):17-35.

Blaxter M, Mieszkowska N, Di Palma F, Holland P, Durbin R, Richards T, Berriman M, Kersey P, Hollingsworth P, Wilson W et al. 2022. Sequence locally, think globally: The darwin tree of life project. Proceedings of the National Academy of Sciences. 119(4):e2115642118.

Boudet A-M. 1998. A new view of lignification. Trends in plant science. 3(2):67-71.

Brunecky R, Alahuhta M, Sammond DW, Xu Q, Chen M, Wilson DB, Brady JW, Himmel ME, Bomble YJ, Lunin VV. 2017. Natural diversity of glycoside hydrolase family 48 exoglucanases: Insights from structure. Biotechnology for biofuels. 10(1):1-9.

Burton RA, Gidley MJ, Fincher GB. 2010. Heterogeneity in the chemistry, structure and function of plant cell walls. Nature chemical biology. 6(10):724-732.

Busch A, Danchin EG, Pauchet Y. 2019. Functional diversification of horizontally acquired glycoside hydrolase family 45 (gh45) proteins in phytophaga beetles. BMC evolutionary biology. 19(1):100.

Busch A, Kunert G, Heckel DG, Pauchet Y. 2017. Evolution and functional characterization of cazymes belonging to subfamily 10 of glycoside hydrolase family 5 (gh5_10) in two species of phytophagous beetles. PloS one. 12(8):e0184305.

Busch A, Kunert G, Wielsch N, Pauchet Y. 2018. Cellulose degradation in gastrophysa viridula (coleoptera: Chrysomelidae): Functional characterization of two cazymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity. Insect molecular biology. 27(5):633-650.

Calderón-Cortés N, Quesada M, Watanabe H, Cano-Camacho H, Oyama K. 2012. Endogenous plant cell wall digestion: A key mechanism in insect evolution. Annual Review of Ecology, Evolution, and Systematics. 43:45-71.

Calderón-Cortés N, Watanabe H, Cano-Camacho H, Zavala-Páramo G, Quesada M. 2010. Cdna cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle oncideres albomarginata chamela (cerambycidae). Insect molecular biology. 19(3):323-336.

Cameron SL, Sullivan J, Song H, Miller KB, Whiting MF. 2009. A mitochondrial genome phylogeny of the neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zoologica scripta. 38(6):575-590.

Cartmell A, McKee LS, Peña MJ, Larsbrink J, Brumer H, Kaneko S, Ichinose H, Lewis RJ, Viksø-Nielsen A, Gilbert HJ. 2011. The structure and function of an arabinan-specific α-1, 2-arabinofuranosidase identified from screening the activities of bacterial gh43 glycoside hydrolases. Journal of Biological Chemistry. 286(17):15483-15495.

Castillo-González A, Burrola-Barraza M, Domínguez-Viveros J, Chávez-Martínez A. 2014. Rumen microorganisms and fermentation. Archivos de Medicina Veterinaria. 46(3):349-361.

Chang C-J, Lee C-C, Chan Y-T, Trudeau DL, Wu M-H, Tsai C-H, Yu S-M, Ho T-HD, Wang AH-J, Hsiao C-D. 2016. Exploring the mechanism responsible for cellulase thermostability by structure-guided recombination. PLoS One. 11(3):e0147485.

Chang C-J, Wu CP, Lu S-C, Chao A-L, Ho T-HD, Yu S-M, Chao Y-C. 2012. A novel exo-cellulase from white spotted longhorn beetle (anoplophora malasiaca). Insect biochemistry and molecular biology. 42(9):629-636.

Chen Z, Erickson DL, Meng J. 2021. Polishing the oxford nanopore long-read assemblies of bacterial pathogens with illumina short reads to improve genomic analyses. Genomics. 113(3):1366-1377.
Cocquempot C, Lindelöw A. 2010. Longhorn beetles (coleoptera, cerambycidae). Pensoft Publishers.
Collins T, Gerday C, Feller G. 2005. Xylanases, xylanase families and extremophilic xylanases. FEMS microbiology reviews. 29(1):3-23.
Cosgrove D, Jarvis m. 2012. Comparative structure and biomechanics of plant primary and secondary cell walls. Frontiers in Plant Science. 3.
Cummings J, Stephen A. 2007. Carbohydrate terminology and classification. European journal of clinical nutrition. 61(1):S5-S18.
Darwin C. 1868. The variation of animals and plants under domestication. J. murray.
Davies G, Henrissat B. 1995. Structures and mechanisms of glycosyl hydrolases. Structure. 3(9):853-859.
Davies GJ, Tolley SP, Henrissat B, Hjort C, Schulein M. 1995. Structures of oligosaccharide-bound forms of the endoglucanase v from humicola insolens at 1.9. Ang. Resolution. Biochemistry. 34(49):16210-16220.
Davison A, Blaxter M. 2005. Ancient origin of glycosyl hydrolase family 9 cellulase genes. Molecular Biology and Evolution. 22(5):1273-1284.
de la Paz Celorio-Mancera M, Allen ML, Powell AL, Ahmadi H, Salemi MR, Phinney BS, Shackel KA, Greve LC, Teuber LR, Labavitch JM. 2008. Polygalacturonase causes lygus-like damage on plants: Cloning and identification of western tarnished plant bug (lygus hesperus) polygalacturonases secreted during feeding. Arthropod-Plant Interactions. 2(4):215-225.
DeBoy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB, Mohamoud Y, Watkins K, Henrissat B, Gilbert HJ. 2008. Insights into plant cell wall degradation from the genome sequence of the soil bacterium cellvibrio japonicus. Journal of bacteriology. 190(15):5455-5463.
Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM. 2002. The draft genome of ciona intestinalis: Insights into chordate and vertebrate origins. Science. 298(5601):2157-2167.
Delalibera I, Jr, Handelsman J, Raffa KF. 2005. Contrasts in cellulolytic activities of gut microorganisms between the wood borer, saperda vestita (coleoptera: Cerambycidae), and the bark beetles, ips pini and dendroctonus frontalis (coleoptera: Curculionidae). Environmental Entomology. 34(3):541-547.
Dhandapani RK, Duan JJ, Palli SR. 2020a. Orally delivered dsrna induces knockdown of target genes and mortality in the asian long-horned beetle, anoplophora glabripennis. Archives of insect biochemistry and physiology. 104(4):e21679.
Dhandapani RK, Gurusamy D, Duan JJ, Palli SR. 2020b. Rnai for management of asian long-horned beetle, anoplophora glabripennis: Identification of target genes. Journal of Pest Science. 93(2):823-832.
Dhawan S, Kaur J. 2007. Microbial mannanases: An overview of production and applications. Critical reviews in biotechnology. 27(4):197-216.
Doblin MS, Pettolino F, Bacic A. 2010. Plant cell walls: The skeleton of the plant world. Functional Plant Biology. 37(5):357-381.
Donev E, Gandla ML, Jönsson LJ, Mellerowicz EJ. 2018. Engineering non-cellulosic polysaccharides of wood for the biorefinery. Frontiers in Plant Science. 9.
Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2021. The carbohydrate-active enzyme database: Functions and literature. Nucleic Acids Research. 50(D1):D571-D577.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP. 2017. De novo assembly of the aedes aegypti genome using hi-c yields chromosome-length scaffolds. Science. 356(6333):92-95.

Edwards M, Dea I, Bulpin P, Reid J. 1986. Purification and properties of a novel xyloglucan-specific endo-(1----4)-beta-d-glucanase from germinated nasturtium seeds (tropaeolum majus l.). Journal of Biological Chemistry. 261(20):9489-9494.

Ellison CE, Cao W. 2019. Nanopore sequencing and hi-c scaffolding provide insight into the evolutionary dynamics of transposable elements and pirna production in wild strains of drosophila melanogaster. Nucleic Acids Research. 48(1):290-303.

Essig EO. 1936. A sketch history of entomology. Osiris. 2:80-123.

Eyun S-i, Wang H, Pauchet Y, Ffrench-Constant RH, Benson AK, Valencia-Jiménez A, Moriyama EN, Siegfried BD. 2014. Molecular evolution of glycoside hydrolase genes in the western corn rootworm (diabrotica virgifera virgifera). PloS one. 9(4):e94052.

Ferreira AH, Marana SR, Terra WR, Ferreira C. 2001. Purification, molecular cloning, and properties of a β-glycosidase isolated from midgut lumen of tenebrio molitor (coleoptera) larvae. Insect Biochemistry and Molecular Biology. 31(11):1065-1076.

Fry SC. 2004. Primary cell wall metabolism: Tracking the careers of wall polymers in living plant cells. New phytologist. 161(3):641-675.

Futuyma DJ, Moreno G. 1988. The evolution of ecological specialization. Annual review of Ecology and Systematics.207-233.

Garron M-L, Cygler M. 2010. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology. 20(12):1547-1573.

Geib SM, Filley TR, Hatcher PG, Hoover K, Carlson JE, Jimenez-Gasco MdM, Nakagawa-Izumi A, Sleighter RL, Tien M. 2008. Lignin degradation in wood-feeding insects. Proceedings of the National Academy of Sciences. 105(35):12932-12937.

Geib SM, Jimenez-Gasco MdM, Carlson JE, Tien M, Hoover K. 2009. Effect of host tree species on cellulase activity and bacterial community composition in the gut of larval asian longhorned beetle. Environmental Entomology. 38(3):686-699.

Geib SM, Tien M, Hoover K. 2010. Identification of proteins involved in lignocellulose degradation using in gel zymogram analysis combined with mass spectroscopy-based peptide analysis of gut proteins from larval asian longhorned beetles, anoplophora glabripennis. Insect Science. 17(3):253-264.

Gilbert HJ, Hazlewood GP. 1993. Bacterial cellulases and xylanases. Microbiology. 139(2):187-194.

Girard C, Jouanin L. 1999a. Molecular cloning of a gut-specific chitinase cdna from the beetle phaedon cochleariae. Insect biochemistry and molecular biology. 29(6):549-556.

Girard C, Jouanin L. 1999b. Molecular cloning of cdnas encoding a range of digestive enzymes from a phytophagous beetle, phaedon cochleariae. Insect Biochemistry and Molecular Biology. 29(12):1129-1142.

Gloster TM, Ibatullin FM, Macauley K, Eklof JM, Roberts S, Turkenburg JP, Bjørnvad ME, Jørgensen PL, Danielsen S, Johansen KS. 2007. Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families gh5 and gh12. Journal of Biological Chemistry. 282(26):19177-19189.

Haack RA. 2017. Feeding biology of cerambycids. In: Wang, Qiao, ed Cerambycidae of the world; biology and pest management Boca Raton, FLP CRC Press: 105-124.105-124.

Haddad S, Shin S, Lemmon AR, Lemmon EM, Svacha P, Farrell B, Ślipinski A, Donald W, D. MD. 2018. Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (cerambycidae). Systematic Entomology. 43(1):68-89.

Hanks LM. 1999. Influence of the larval host plant on reproductive strategies of cerambycid beetles. Annual review of entomology. 44(1):483-505.
Hasper AA, Dekkers E, van Mil M, van de Vondervoort PJ, de Graaff LH. 2002. Eglc, a new endoglucanase from aspergillus niger with major activity towards xyloglucan. Applied and Environmental Microbiology. 68(4):1556-1560.
He S, Jiang B, Chakraborty A, Yu G. 2022. The evolution of glycoside hydrolase family 1 in insects related to their adaptation to plant utilization. Insects. 13(9):786.
Hearle JWS, Peters RH. 2013. Fibre structure. Elsevier.
Heinonen E, Henriksson G, Lindström ME, Vilaplana F, Wohlert J. 2022. Xylan adsorption on cellulose: Preferred alignment and local surface immobilizing effect. Carbohydrate Polymers. 285:119221.
Heitz E. 1927. Über intrazelluläre symbiose bei holzfressenden käferlarven i. Zeitschrift für Morphologie und Ökologie der Tiere.279-305.
Hemsworth GR, Henrissat B, Davies GJ, Walton PH. 2014. Discovery and characterization of a new family of lytic polysaccharide monooxygenases. Nature chemical biology. 10(2):122-126.
Henrissat B, Claeyssens M, Tomme P, Lemesle L, Mornon J-P. 1989. Cellulase families revealed by hydrophobic cluster analysi. Gene. 81(1):83-95.
Hilge M, Gloor SM, Rypniewski W, Sauer O, Heightman TD, Zimmermann W, Winterhalter K, Piontek K. 1998. High-resolution native and complex structures of thermostable β-mannanase from thermomonospora fusca–substrate specificity in glycosyl hydrolase family 5. Structure. 6(11):1433-1444.
Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. science. 315(5813):804-807.
Hotaling S, Sproul JS, Heckenhauer J, Powell A, Larracuente AM, Pauls SU, Kelley JL, Frandsen PB. 2021. Long reads are revolutionizing 20 years of insect genome sequencing. Genome Biology and Evolution. 13(8).
Ibarra LN, Alves AEOdA, Antonino JD, Prado GS, Pinto CEM, Soccol CR, Vasconcelos ÉARd, Grossi-de-Sa MF. 2019. Enzymatic activity of a recombinant β-1, 4-endoglucanase from the cotton boll weevil (anthonomus grandis) aiming second generation ethanol production. Scientific Reports. 9(1):1-10.
Innan H, Kondrashov F. 2010. The evolution of gene duplications: Classifying and distinguishing between models. Nature Reviews Genetics. 11(2):97-108.
Ishida T, Yaoi K, Hiyoshi A, Igarashi K, Samejima M. 2007. Substrate recognition by glycoside hydrolase family 74 xyloglucanase from the basidiomycete phanerochaete chrysosporium. The FEBS Journal. 274(21):5727-5736.
Jayani RS, Saxena S, Gupta R. 2005. Microbial pectinolytic enzymes: A review. Process Biochemistry. 40(9):2931-2944.
Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, Allen JM, Petersen M, Donath A, Walden KKO et al. 2018. Phylogenomics and the evolution of hemipteroid insects. Proceedings of the National Academy of Sciences. 115(50):12775-12780.
Jones A, Winge P, Bones A, Cole R, Rossiter J. 2002. Characterization and evolution of a myrosinase from the cabbage aphid brevicoryne brassicae. Insect biochemistry and molecular biology. 32(3):275-284.
Kang X, Kirui A, Dickwella Widanage MC, Mentink-Vigier F, Cosgrove DJ, Wang T. 2019. Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state nmr. Nature communications. 10(1):1-9.
Katahira R, Elder TJ, Beckham GT. 2018. A brief introduction to lignin structure.

Kaur P, Bhardwaj NK, Sharma J. 2016. Pretreatment with xylanase and its significance in hemicellulose removal from mixed hardwood kraft pulp as a process step for viscose. Carbohydrate polymers. 145:95-102.
Keena MA. 2017. Laboratory rearing and handling of cerambycids. In: Wang, Qiao, ed Cerambycidae of the world: biology and pest management Boca Raton, FL: CRC Press: 253-289 Chapter 7.253-289.
Ketudat Cairns JR, Esen A. 2010. B-glucosidases. Cellular and Molecular Life Sciences. 67(20):3389-3405.
Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N. 2004. A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode bursaphelenchus xylophilus. FEBS letters. 572(1-3):201-205.
King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ, Bruce NC, Graham IA, McQueen-Mason SJ. 2010. Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. Proceedings of the National Academy of Sciences. 107(12):5345-5350.
Kirsch R, Gramzow L, Theißen G, Siegfried BD, Ffrench-Constant R, 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 biochemistry and molecular biology. 52:33-50.
Kirsch R, Kunert G, Vogel H, Pauchet Y. 2019. Pectin digestion in herbivorous beetles: Impact of pseudoenzymes exceeds that of their active counterparts. Frontiers in physiology. 10:685.
Kirsch R, Wielsch N, Vogel H, Svatoš A, Heckel DG, Pauchet Y. 2012. Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. Bmc Genomics. 13(1):1-15.
Kirsch RO, Yu; Haeger, Wiebke; Vogel, Heiko; Kunert, Grit; Pauchet, Yannick. 2022. Metabolic novelty originating from horizontal gene transfer is essential for leaf beetle survival. PNAS. In Press.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nature biotechnology. 37(5):540-546.
Koonin EV. 2009. Darwinian evolution in the light of genomics. Nucleic Acids Research. 37(4):1011-1034.
Koshland Jr D. 1953. Stereochemistry and the mechanism of enzymatic reactions. Biological reviews. 28(4):416-436.
Kukor JJ, Cowan DP, Martin MM. 1988. The role of ingested fungal enzymes in cellulose digestion in the larvae of cerambycid beetles. Physiological Zoology. 61(4):364-371.
Kunieda T, Fujiyuki T, Kucharski R, Foret S, Ament S, Toth A, Ohashi K, Takeuchi H, Kamikouchi A, Kage E. 2006. Carbohydrate metabolism genes and pathways in insects: Insights from the honey bee genome. Insect molecular biology. 15(5):563-576.
Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: Structures, functions, and mechanisms. Annual Review of Biochemistry. 77(1):521-555.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409(6822):860-921.
Larsson AM, Anderson L, Xu B, Muñoz IG, Usón I, Janson J-C, Stålbrand H, Ståhlberg J. 2006. Three-dimensional crystal structure and enzymic characterization of β-mannanase man5a from blue mussel mytilus edulis. Journal of molecular biology. 357(5):1500-1510.

Lee SJ, Kim SR, Yoon HJ, Kim I, Lee KS, Je YH, Lee SM, Seo SJ, Sohn HD, Jin BR. 2004. Cdna cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, apriona germari. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 139(1):107-116.

Lee SJ, Lee KS, Kim SR, Gui ZZ, Kim YS, Yoon HJ, Kim I, Kang PD, Sohn HD, Jin BR. 2005. A novel cellulase gene from the mulberry longicorn beetle, apriona germari: Gene structure, expression, and enzymatic activity. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 140(4):551-560.

Lee TM, Farrow MF, Arnold FH, Mayo SL. 2011. A structural study of hypocrea jecorina cel5a. Protein Science. 20(11):1935-1940.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the cazy database to integrate auxiliary redox enzymes. Biotechnology for Biofuels. 6(1):41.

Li F, Zhao X, Li M, He K, Huang C, Zhou Y, Li Z, Walters JR. 2019. Insect genomes: Progress and challenges. Insect molecular biology. 28(6):739-758.

Lin Y-H, McLenachan PA, Gore AR, Phillips MJ, Ota R, Hendy MD, Penny D. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Molecular Biology and Evolution. 19(12):2060-2070.

Linsley EG. 1959. Ecology of cerambycidae. Annual review of entomology. 4(1):99-138.

Linsley EG, Chemsak JA. 1961. The cerambycidae of north america. Univ of California Press.

Liu J, Wang X, Xu D. 2010. Qm/mm study on the catalytic mechanism of cellulose hydrolysis catalyzed by cellulase cel5a from acidothermus cellulolyticus. The Journal of Physical Chemistry B. 114(3):1462-1470.

Lu H, Giordano F, Ning Z. 2016. Oxford nanopore minion sequencing and genome assembly. Genomics, proteomics & bioinformatics. 14(5):265-279.

Malgas S, van Dyk JS, Pletschke BI. 2015. A review of the enzymatic hydrolysis of mannans and synergistic interactions between β-mannanase, β-mannosidase and α-galactosidase. World journal of microbiology and biotechnology. 31(8):1167-1175.

Marana SR, Jacobs-Lorena M, Terra WR, Ferreira C. 2001. Amino acid residues involved in substrate binding and catalysis in an insect digestive β-glycosidase. Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology. 1545(1-2):41-52.

Marques AR, Coutinho PM, Videira P, Fialho AM, Sá-Correia I. 2003. Sphingomonas paucimobilis beta-glucosidase bgl1: A member of a new bacterial subfamily in glycoside hydrolase family 1. Biochemical Journal. 370(3):793-804.

Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS biology. 9(12):e1001221.

Martin MM. 1983. Cellulose digestion in insects. Comparative Biochemistry and Physiology Part A: Physiology. 75(3):313-324.

Martin MM. 1991. The evolution of cellulose digestion in insects. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 333(1267):281-288.

Martinez-Fleites C, Guerreiro CI, Baumann MJ, Taylor EJ, Prates JA, Ferreira LM, Fontes CM, Brumer H, Davies GJ. 2006. Crystal structures of clostridium thermocellum xyloglucanase, xgh74a, reveal the structural basis for xyloglucan recognition and degradation. Journal of Biological Chemistry. 281(34):24922-24933.

Mattson Jr WJ. 1980. Herbivory in relation to plant nitrogen content. Annual review of ecology and systematics. 11(1):119-161.

McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn S-J, Arsala D. 2016. Genome of the asian longhorned beetle (anoplophora glabripennis), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. Genome biology. 17(1):1-18.

McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE, Friedrich F, Letsch H. 2019. The evolution and genomic basis of beetle diversity. Proceedings of the National Academy of Sciences. 116(49):24729-24737.

Mellerowicz EJ, Sundberg B. 2008. Wood cell walls: Biosynthesis, developmental dynamics and their implications for wood properties. Current Opinion in Plant Biology. 11(3):293-300.

Mendel G. 1866. Versuche uber pflanzen-hybriden. Verhandlungen des naturforschenden Vereins in Brunn fur. 4:3-47.

Meng P, Hoover K, Keena M. 2015. Asian longhorned beetle (coleoptera: Cerambycidae), an introduced pest of maple and other hardwood trees in north america and europe. Journal of Integrated Pest Management. 6(1).

Mohnen D. 2008. Pectin structure and biosynthesis. Current opinion in plant biology. 11(3):266-277.

Monné M, Wang Q. 2017. General morphology, classification and biology of cerambycidae. CRC Press.

Montanier C, Van Bueren AL, Dumon C, Flint JE, Correia MA, Prates JA, Firbank SJ, Lewis RJ, Grondin GG, Ghinet MG. 2009. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proceedings of the National Academy of Sciences. 106(9):3065-3070.

Naumoff DG. 2001. B-fructosidase superfamily: Homology with some α-l-arabinases and β-d-xylosidases. Proteins: Structure, Function, and Bioinformatics. 42(1):66-76.

O'neill M, Albersheim P, Darvill A. 1990. The pectic polysaccharides of primary cell walls. Methods in plant biochemistry. Elsevier. p. 415-441.

Okmane L, Nestor G, Jakobsson E, Xu B, Igarashi K, Sandgren M, Kleywegt GJ, Ståhlberg J. 2022. Glucomannan and beta-glucan degradation by mytilus edulis cel45a: Crystal structure and activity comparison with gh45 subfamily a, b and c. Carbohydrate Polymers. 277:118771.

Padilla-Hurtado B, Flórez-Ramos C, Aguilera-Gálvez C, Medina-Olaya J, Ramírez-Sanjuan A, Rubio-Gómez J, Acuña-Zornosa R. 2012. Cloning and expression of an endo-1, 4-β-xylanase from the coffee berry borer, hypothenemus hampei. BMC research notes. 5(1):23.

Palomares-Rius JE, Hirooka Y, Tsai IJ, Masuya H, Hino A, Kanzaki N, Jones JT, Kikuchi T. 2014. Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. BMC evolutionary biology. 14(1):1-12.

Park YB, Cosgrove DJ. 2015. Xyloglucan and its interactions with other components of the growing cell wall. Plant and Cell Physiology. 56(2):180-194.

Pauchet Y, Heckel DG. 2013. The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer. Proceedings of the Royal Society B: Biological Sciences. 280(1763):20131021.

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 biochemistry and molecular biology. 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(10):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(12):e15635.
Pauly M, Keegstra K. 2016. Biosynthesis of the plant cell wall matrix polysaccharide xyloglucan. Annual review of plant biology. 67:235-259.
Poelchau M, Childers C, Moore G, Tsavatapalli V, Evans J, Lee C-Y, Lin H, Lin J-W, Hackett K. 2015. The i5k workspace@ nal—enabling genomic data access, visualization and curation of arthropod genomes. Nucleic acids research. 43(D1):D714-D719.
Prajapati AS, Panchal KJ, Pawar VA, Noronha MJ, Patel DH, Subramanian R. 2018. Review on cellulase and xylanase engineering for biofuel production. Industrial Biotechnology. 14(1):38-44.
Puchart V, Vršanská M, Svoboda P, Pohl J, Ögel ZB, Biely P. 2004. Purification and characterization of two forms of endo-β-1, 4-mannanase from a thermotolerant fungus, aspergillus fumigatus imi 385708 (formerly thermomyces lanuginosus imi 158749). Biochimica et Biophysica Acta (BBA)-General Subjects. 1674(3):239-250.
Raje KR, Ferris VR, Holland JD. 2016. Phylogenetic signal and potential for invasiveness. Agricultural and Forest Entomology. 18(3):260-269.
Rexová-Benková Ľ, Markovič O. 1976. Pectic enzymes. Advances in carbohydrate chemistry and biochemistry. 33:323-385.
Reynolds S. 2019. Cooking up the perfect insect: Aristotle's transformational idea about the complete metamorphosis of insects. Philosophical Transactions of the Royal Society B. 374(1783):20190074.
Ribeiro AJ, Tyzack JD, Borkakoti N, Holliday GL, Thornton JM. 2020. A global analysis of function and conservation of catalytic residues in enzymes. Journal of Biological Chemistry. 295(2):314-324.
Ridley BL, O'Neill MA, Mohnen D. 2001. Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry. 57(6):929-967.
Robertson HM. 1993. The mariner transposable element is widespread in insects. Nature. 362(6417):241-245.
Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 425(6960):798-804.
Rongpipi S, Ye D, Gomez ED, Gomez EW. 2019. Progress and opportunities in the characterization of cellulose – an important regulator of cell wall growth and mechanics. Frontiers in Plant Science. 9.
Sakamoto K, Toyohara H. 2009. Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam corbicula japonica. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 152(4):390-396.
Salmon M, Thimmappa RB, Minto RE, Melton RE, Hughes RK, O'Maille PE, Hemmings AM, Osbourn A. 2016. A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. Proceedings of the National Academy of Sciences. 113(30):E4407-E4414.
Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, Hutchison C, Slocombe PM, Smith M. 1977. Nucleotide sequence of bacteriophage φx174 DNA. nature. 265(5596):687-695.
Scharf ME, Kovaleva ES, Jadhao S, Campbell JH, Buchman GW, Boucias DG. 2010. Functional and translational analyses of a beta-glucosidase gene (glycosyl hydrolase

family 1) isolated from the gut of the lower termite reticulitermes flavipes. Insect biochemistry and molecular biology. 40(8):611-620.
Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annual review of plant biology. 61:263-289.
Schloss PD, Delalibera Jr I, Handelsman J, Raffa KF. 2006. Bacteria associated with the guts of two wood-boring beetles: Anoplophora glabripennis and saperda vestita (cerambycidae). Environmental Entomology. 35(3):625-629.
Schomann H. 1937. Die symbiose der bockkäfer. Zeitschrift für Morphologie und Ökologie der Tiere.542-612.
Schröder R, Wegrzyn TF, Sharma NN, Atkinson RG. 2006. Leman4 endo-β-mannanase from ripe tomato fruit can act as a mannan transglycosylase or hydrolase. Planta. 224(5):1091-1102.
Scrivener A, Watanabe H, Noda H. 1997. Diet and carbohydrate digestion in the yellow-spotted longicorn beetle psacothea hilaris. Journal of insect physiology. 43(11):1039-1052.
Scully ED, Hoover K, Carlson JE, Tien M, Geib SM. 2013. Midgut transcriptome profiling of anoplophora glabripennis, a lignocellulose degrading cerambycid beetle. BMC genomics. 14(1):1-26.
Seiboth B, Messner R, Gruber F, Kubicek CP. 1992. Disruption of the trichoderma reesei cbh2 gene coding for cellobiohydrolase ii leads to a delay in the triggering of cellulase formation by cellulose. Microbiology. 138(6):1259-1264.
Sheffield N, Song H, Cameron S, Whiting M. 2008. A comparative analysis of mitochondrial genomes in coleoptera (arthropoda: Insecta) and genome descriptions of six new beetles. Molecular biology and evolution. 25(11):2499-2509.
Shelomi M, Heckel DG, Pauchet Y. 2016. Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (phasmatodea). Insect Biochemistry and Molecular Biology. 71:1-11.
Shelomi M, Watanabe H, Arakawa G. 2014. Endogenous cellulase enzymes in the stick insect (phasmatodea) gut. Journal of insect physiology. 60:25-30.
Shen Z, Denton M, Mutti N, Pappan K, Michael R K, Reese JC, R Reeck G. 2003. Polygalacturonase from sitophilus oryzae: Possible horizontal transfer of a pectinase gene from fungi to weevils. Journal of Insect Science. 3(1).
Sheppard PO, Grant FJ, Oort PJ, Sprecher CA, Foster DC, Hagen FS, Upshall A, McKnight GL, O'Hara PJ. 1994. The use of conserved cellulase family-specific sequences to clone cellulase homologue cdnas from fusarium oxysporum. Gene. 150(1):163-167.
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(4):784-797.
Shoham Y, Schwartz Z, Khasin A, Gat O, Zosim Z, Rosenberg E. 1992. Delignification of wood pulp by a thermostable xylanase from bacillus stearothermophilus strain t-6. Microorganisms to combat pollution. Springer. p. 83-94.
Shrinivas D, Savitha G, Raviranjan K, Naik GR. 2010. A highly thermostable alkaline cellulase-free xylanase from thermoalkalophilic bacillus sp. Jb 99 suitable for paper and pulp industry: Purification and characterization. Applied biochemistry and biotechnology. 162(7):2049-2057.
Smith MJ, Turgeon JJ, De Groot P, Gasman B. 2009. Asian longhorned beetle anoplophora glabripennis (motschulsky): Lessons learned and opportunities to improve the process of eradication and management. American Entomologist. 55(1):21.

Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T. 2004. Toward a systems approach to understanding plant cell walls. Science. 306(5705):2206-2211.

Song H, Sheffield NC, Cameron SL, Miller KB, Whiting MF. 2010. When phylogenetic assumptions are violated: Base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. Systematic Entomology. 35(3):429-448.

Song JM, Hong SK, An YJ, Kang MH, Hong KH, Lee Y-H, Cha S-S. 2017. Genetic and structural characterization of a thermo-tolerant, cold-active, and acidic endo-β-1, 4-glucanase from antarctic springtail, cryptopygus antarcticus. Journal of agricultural and food chemistry. 65(8):1630-1640.

Song JM, Nam K-W, Kang SG, Kim C-G, Kwon S-T, Lee Y-H. 2008. Molecular cloning and characterization of a novel cold-active β-1, 4-d-mannanase from the antarctic springtail, cryptopygus antarcticus. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 151(1):32-40.

Steenbakkers PJ, Ubhayasekera W, Goossen HJ, van LIEROP EM, van der DRIFT C, Vogels GD, Mowbray SL, den Camp HJO. 2002. An intron-containing glycoside hydrolase family 9 cellulase gene encodes the dominant 90 kda component of the cellulosome of the anaerobic fungus piromyces sp. Strain e2. Biochemical Journal. 365(1):193-204.

Stern R, Jedrzejas MJ. 2008. Carbohydrate polymers at the center of life's origins: The importance of molecular processivity. Chemical reviews. 108(12):5061-5085.

Sturtevant AH. 1913. The linear arrangement of six sex? Linked factors in drosophila, as shown by their mode of association. Journal of experimental zoology. 14(1):43-59.

Subramaniyan S, Prema P. 2002. Biotechnology of microbial xylanases: Enzymology, molecular biology, and application. Critical reviews in biotechnology. 22(1):33-64.

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(16):3455-3460.

Suzuki Ki, Ojima T, Nishita K. 2003. Purification and cdna cloning of a cellulase from abalone haliotis discus hannai. European Journal of Biochemistry. 270(4):771-778.

Suzuki S, Fukuoka M, Ookuchi H, Sano M, Ozeki K, Nagayoshi E, Takii Y, Matsushita M, Tada S, Kusumoto K-I. 2010. Characterization of aspergillus oryzae glycoside hydrolase family 43 β-xylosidase expressed in escherichia coli. Journal of bioscience and bioengineering. 109(2):115-117.

Švácha P, Lawrence J. 2014. 2.1 vesperidae mulsant, 1839; 2.2 oxypeltidae lacordaire, 1868; 2.3 disteniidae j. Thomson, 1861; 2.4 cerambycidae latreille, 1802. Handbook of zoology, Arthropoda: insecta.16-177.

Talbot G, Sygusch J. 1990. Purification and characterization of thermostable beta-mannanase and alpha-galactosidase from bacillus stearothermophilus. Applied and Environmental Microbiology. 56(11):3505-3510.

Teeri TT. 1997. Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. Trends in biotechnology. 15(5):160-167.

Thapa S, Mishra J, Arora N, Mishra P, Li H, Bhatti S, Zhou S. 2020. Microbial cellulolytic enzymes: Diversity and biotechnology with reference to lignocellulosic biomass degradation. Reviews in Environmental Science and Bio/Technology. 19(3):621-648.

Thibault J-F, Ralet M-C. 2008. 32 pectins, their origin, structure and functions. Advanced dietary fibre technology.369-378.

Tokuda G. 2019. Plant cell wall degradation in insects: Recent progress on endogenous enzymes revealed by multi-omics technologies. Advances in Insect Physiology. 57:97-136.

Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK. 2012. Advances in insect phylogeny at the dawn of the postgenomic era. Annual review of entomology. 57(1):449-468.
Urbanowicz BR, Catalá C, Irwin D, Wilson DB, Ripoll DR, Rose JKC. 2007. A tomato endo-β-1,4-glucanase, slcel9c1, represents a distinct subclass with a new family of carbohydrate binding modules (cbm49)*. Journal of Biological Chemistry. 282(16):12066-12074.
Valencia A, Alves AP, Siegfried BD. 2013. Molecular cloning and functional characterization of an endogenous endoglucanase belonging to ghf45 from the western corn rootworm, diabrotica virgifera virgifera. Gene. 513(2):260-267.
Valls A, Diaz P, Pastor F, Valenzuela SV. 2016. A newly discovered arabinoxylan-specific arabinofuranohydrolase. Synergistic action with xylanases from different glycosyl hydrolase families. Applied microbiology and biotechnology. 100(4):1743-1751.
Vanholme R, De Meester B, Ralph J, Boerjan W. 2019. Lignin biosynthesis and its integration into metabolism. Current Opinion in Biotechnology. 56:230-239.
Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. 2010. Lignin biosynthesis and structure. Plant physiology. 153(3):895-905.
Viikari L, Kantelinen A, Sundquist J, Linko M. 1994. Xylanases in bleaching: From an idea to the industry. FEMS Microbiology Reviews. 13(2-3):335-350.
Vorwerk S, Somerville S, Somerville C. 2004. The role of plant cell wall polysaccharide composition in disease resistance. Trends in plant science. 9(4):203-209.
Wang F, Li D, Wang Z, Dong A, Liu L, Wang B, Chen Q, Liu X. 2014. Transcriptomic analysis of the rice white tip nematode, aphelenchoides besseyi (nematoda: Aphelenchoididae). PloS one. 9(3):e91591.
Wang Q. 2017. Cerambycidae of the world: Biology and pest management. CRC press.
Watanabe H, Noda H, Tokuda G, Lo N. 1998. A cellulase gene of termite origin. Nature. 394(6691):330-331.
Watanabe H, Tokuda G. 2001. Animal cellulases. Cellular and Molecular Life Sciences CMLS. 58(9):1167-1178.
Watanabe H, Tokuda G. 2010. Cellulolytic systems in insects. Annual review of entomology. 55.
Watson JD, Crick FH. 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature. 171(4356):737-738.
Weimer PJ. 1992. Cellulose degradation by ruminal microorganisms. Critical Reviews in Biotechnology. 12(3):189-223.
Willow J, Veromann E. 2021. Highly variable dietary rnai sensitivity among coleoptera. Frontiers in Plant Science.2914.
Wilson DB. 2009. Cellulases and biofuels. Current opinion in biotechnology. 20(3):295-299.
Wolf S, Mouille G, Pelloux J. 2009. Homogalacturonan methyl-esterification and plant development. Molecular plant. 2(5):851-860.
Wu G-L, Kuo T-H, Tsay T-T, Tsai IJ, Chen PJ. 2016a. Glycoside hydrolase (gh) 45 and 5 candidate cellulases in aphelenchoides besseyi isolated from bird's-nest fern. PloS one. 11(7):e0158663.
Wu J, Wang Y, Park S-Y, Kim SG, Yoo JS, Park S, Gupta R, Kang KY, Kim ST. 2016b. Secreted alpha-n-arabinofuranosidase b protein is required for the full virulence of magnaporthe oryzae and triggers host defences. PLoS One. 11(10):e0165149.
Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal gene transfer contributes to the evolution of arthropod herbivory. Genome biology and evolution. 8(6):1785-1801.

Xiao C, Anderson C. 2013. Roles of pectin in biomass yield and processing for biofuels. Frontiers in Plant Science. 4.

Xue H-J, Niu Y-W, Segraves KA, Nie R-E, Hao Y-J, Zhang L-L, Cheng X-C, Zhang X-W, Li W-Z, Chen R-S. 2021. The draft genome of the specialist flea beetle altica viridicyanea (coleoptera: Chrysomelidae). BMC genomics. 22(1):1-18.

Yadav S, Yadav PK, Yadav D, Yadav KDS. 2009. Pectin lyase: A review. Process Biochemistry. 44(1):1-10.

Yang X, Shi P, Ma R, Luo H, Huang H, Yang P, Yao B. 2015. A new gh43 α-arabinofuranosidase from humicola insolens y1: Biochemical characterization and synergistic action with a xylanase on xylan degradation. Applied biochemistry and biotechnology. 175(4):1960-1970.

Yaoi K, Kondo H, Hiyoshi A, Noro N, Sugimoto H, Tsuda S, Mitsuishi Y, Miyazaki K. 2007. The structural basis for the exo-mode of action in gh74 oligoxyloglucan reducing end-specific cellobiohydrolase. Journal of molecular biology. 370(1):53-62.

Yaoi K, Nakai T, Kameda Y, Hiyoshi A, Mitsuishi Y. 2005. Cloning and characterization of two xyloglucanases from paenibacillus sp. Strain km21. Applied and environmental microbiology. 71(12):7670-7678.

Yuan JS, Yang X, Lai J, Lin H, Cheng Z-M, Nonogaki H, Chen F. 2007. The endo-β-mannanase gene families in arabidopsis, rice, and poplar. Functional & integrative genomics. 7(1):1-16.

Yuan Y, Bayer PE, Batley J, Edwards D. 2017. Improvements in genomic technologies: Application to crop genomics. Trends in Biotechnology. 35(6):547-558.

Zhao C, Doucet D, Mittapalli O. 2014. Characterization of horizontally transferred β-fructofuranosidase (scrb) genes in a grilus planipennis. Insect molecular biology. 23(6):821-832.

Zhao Q. 2016. Lignification: Flexibility, biosynthesis and regulation. Trends in plant science. 21(8):713-721.

Zhao Y, Man Y, Wen J, Guo Y, Lin J. 2019. Advances in imaging plant cell walls. Trends in plant science. 24(9):867-878.

Zhu Y, Han L, Hefferon KL, Silvaggi NR, Wilson DB, McBride MJ. 2016. Periplasmic cytophaga hutchinsonii endoglucanases are required for use of crystalline cellulose as the sole source of carbon and energy. Applied and environmental microbiology. 82(15):4835-4845.

A. Detailed Author Contributions

i. Manuscript 1

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 D. McKenna, Yannick Pauchet

Systematic Entomology (2021), 46 (4), pp. 784 - 797 (2021), doi: 10.1111/syen.12488

Author contribution	Conceptualization: **NRS (30%)**, DDM, YP Collected specimens: PS, OD, SA Designed experiments: **NRS (70%)**, SGS, YO, BH, DDM, YP Performed experiments: **NRS (70%)**, SGS, VL, YP Data analysis: **NRS (70%)**, SGS, VL, RK, BH, DDM, YP Data visualization: **NRS (75%)**, YP Writing – original draft: **NRS (80%)**, YP Writing – review and editing: **NRS (50%)**, PS, DDM, YP

The project was built on fundamental studies and ideas from YP. NRS, DDM, and YP conceptualized the initial project. Cerambycid larvae were collected by PS, OD, and SA (Table 1). YP performed RNA extraction and sequencing, and NRS synthesized cDNA libraries. Transcriptomes were assembled and analyzed by NRS. SGS and DDM inferred the first draft of a species phylogeny using transcriptomes. After generating scripts to infer species phylogeny by YO, NRS performed to build species phylogeny as the final version (Figure 1). VL and BH contributed to screening transcriptomes to detect putative genes encoding CAZymes using the CAZy pipeline. NRS, RK, and YP designed to curate genes encoding PCWDEs using BLAST tools and RACE PCRs in cerambycid beetles (Figure 2). NRS performed RACE-PCR and cloning to annotate the detected PCWDE-encoding genes, including ORFs. NRS and YP designed to identify the evolutionary relationships of cerambycid-derived PCDWEs with microorganisms. NRS performed all bioinformatics analyses to infer the phylogeny of PCWDEs (Figure 3). Also, YP assembled the genome and analyzed it (Table2, Figure 4). NRS analyzed the data from all experiments with contributions from RK and YP. NRS created most figures and wrote the original draft with the contribution of YP. All authors revised the manuscript; PS, DDM, and YP improved the first draft, and the draft was rewritten into the final manuscript by NRS.

Manuscript No. 1

Short reference Shin et al (2021), Systematic Entomology

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles):

Figure(s) # _1,2,3_*	☑	100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	☐	0% (the data presented in this figure are based exclusively on the work of other co-authors)
	☐	Approximate contribution of the doctoral candidate to the figure: _____% Brief description of the contribution: *(e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)*
** Can refer to more than one fig. if the answer is the same*		

Figure(s) # _4_*	☐	100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	☐	0% (the data presented in this figure are based exclusively on the work of other co-authors)
	☑	Approximate contribution of the doctoral candidate to the figure: _50_% Brief description of the contribution: *(e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)*
** Can refer to more than one fig. if the answer is the same*		

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

ii. Manuscript 2

Duplication of Horizontally Acquired GH5_2 Enzymes Played a Central Role in the Evolution of Longhorned Beetles

Na Ra Shin, Daniel Doucet, Yannick Pauchet

Molecular Biology and Evolution (2022) Volume 39, Issue 6, msac128, doi: 10.1093/molbev/msac128

Author contribution	Conceptualization: **NRS (40%)** , YP Designed experiments: **NRS (50%)**, YP Performed experiments: **NRS (75%)** Data analysis: **NRS (80%)**, YP Data visualization: **NRS (100%)** Writing – original draft: **NRS (80%)**, YP Writing – review and editing: **NRS (65%)**, YP

The project was built on fundamental studies and ideas from YP. NRS and YP conceived the conceptualization of the project. DD provided the materials of *Tetropium fuscum*. YP performed cloning and gene expression for activity assays of GH5_2 in two species (Figure S6-S7). NRS performed cloning and gene expression for activity assays of GH5_2 in seven species (Figure S1-S5). NRS and YP designed all bioinformatics analyses, and NRS performed all bioinformatics analyses and following experiments (Figure 1-6). NRS created all figures and wrote the first draft of the manuscript with a contribution from YP. All authors revised the manuscript, and NRS and YP improved the first draft under review. NRS and YP rewrote the final draft.

Manuscript No. 2

Short reference Shin et al (2022), Molecular Biology and Evolution

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles):

Figure(s) # _1,2,3,4,5,6_ *	☑	100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	☐	0% (the data presented in this figure are based exclusively on the work of other co-authors)
	☐	Approximate contribution of the doctoral candidate to the figure: _____% Brief description of the contribution: *(e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)*
** Can refer to more than one fig. if the answer is the same*		

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

iii. Manuscript 3

Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles

Na Ra Shin, Yu Okamura, Roy Kirsch and Yannick Pauchet

Submitted to Insect Molecular Biology, Status: under reviewed

Author contribution

Conceptualization: **NRS (30%)** , YP
Designed experiments: **NRS (60%)**, YO, RK, YP
Performed experiments: **NRS (70%)** , YP
Data analysis: **NRS (80%)**, YP
Data visualization: **NRS (100%)**
Writing – original draft: **NRS (80%)**, YP
Writing – review and editing: **NRS (60%)**, YO, RK, YP

The project was built on fundamental studies and ideas from RK and YP. NRS, YO, and YP conceived the conceptualization of the project. All authors (RK, NRS, YO, and YP) participated in designing the experiments. RK and NRS extracted gDNA from species and performed Oxford nanopore sequencing with contribution by YP. YO and NR polished and assembled the genome and inferred species phylogeny (Figure 2). NRS performed all bioinformatics analyses on the genome dataset (Figure 3-8). NRS created all figures in the manuscript. NRS and YP wrote the first draft of the manuscript. All authors revised the manuscript and improved the first draft in feedback. NRS and YP rewrote the final manuscript.

Manuscript No. 3

Short reference

Contribution of the doctoral candidate

Contribution of the doctoral candidate to figures reflecting experimental data (only for original articles):

Figure(s) # _1,2,3,4,5,6,7,8__*	☑	100% (the data presented in this figure come entirely from experimental work carried out by the candidate)
	☐	0% (the data presented in this figure are based exclusively on the work of other co-authors)
	☐	Approximate contribution of the doctoral candidate to the figure: _____% Brief description of the contribution: *(e.g. "Figure parts a, d and f" or "Evaluation of the data" etc.)*
** Can refer to more than one fig. if the answer is the same*		

_____ _____

Signature candidate Signature supervisor (member of the Faculty)

B. Acknowledgment

It is a great pleasure to thank the many people who helped me finish my Ph.D. dissertation.

First, I would like to express my special thanks to **Yannick**, a wonderful and inspiring advisor. Without you, I would not have been able to accomplish so many things during my work. I also want to thank you for being patient and helping me grow independently. I also sincerely appreciate the learning opportunity you provided. The experience in your team has been critical in my academic career.

I would also like to thank **David** for allowing me to work as a member of the entomology department. Your feedback and questions have always made me think more deeply about my project and results and helped me to improve. Also, thank you for reviewing this paper.

I also thank **Prof. Dr. Ralf Oelmüller** for being my supervisor from the FSU Jena. Thank you for your comments at my committee meeting and for helping me prepare the paperwork this time. Also, thank you for reviewing this work. I would also like to thank **Prof. Jean-Guy Berrin** for reviewing my thesis.

Thank you, **Martin**, for making me a member of this group. Your questions and encouragement have always been a great help.

I want to thank you, **Heiko**, for allowing me genome sequencing with MinION and giving me always comments, and thank you **Franzi**, for your comments on my work.

Thank you, **Roy** and **Yu**, for being great colleagues who always have great discussions and give helpful advice. It was delightful to discuss science with you.

I would like to thank **Bianca** for helping me be a good scientist in the lab. Without you, I could not have run my best experiment ever. I want to thank **Katrin** for helping me stay in Germany

legally with a visa! I appreciate **Domenica** for sequencing all my PCWDEs encoding genes. With your help, I built my PCWDEs dataset. Also, thank you, **Jette**, for inheriting your knowledge of extracting DNA. I also want to thank you **Steffi** for sharing your amazing python script! I would like to thank **Claudia** for helping all IMPRS doctoral researchers.
Moreover, I would like to thank **Martin** and **Hendrik** for always maintaining a good server and cluster to run pipelines and the rest of the IT department for their support.

I want to express my big thank you to all, **Wiebke**, **Pauline**, **Maike**, and **Olivia**. I appreciate your encouragement and good words to complete my work well. Especially without all your help, I might have been translating the summary into German forever.
Also, thank you all member of **Ent-grp** and **Sym-grp** so much for creating a good atmosphere for fun research.

Megha, **Bhawna** thank you for all five years with you guys. You guys are my first friends in Jena. I thank the **Knitting club**, **Cee**, **Pao**, **Lira**, **Amy**, **Marianna**, and **Lili**. It was really fun to knit many beautiful things together, and it was nice to be able to relax and have a good time after work.

Haewon, big thank you for all your support during my Ph.D time. Also, I really appreciate **Mom** and **Dad**, for all your support mentally and financially during my student period! **Wooju**, my precious another doggy brother, your smile made me keep working hard abroad.

D. Publications and Presentations

Publication

Shin, N. R.; Doucet D.; Pauchet. Y.: *Duplication of horizontally acquired GH5_2 enzymes played a central role in the evolution of longhorned beetles*, Molecular Biology and Evolution, Volume 39, Issue 6, msac128 (2022)

Shin, N. R.; Shin, S.; Okamura, Y.; Kirsch, R.; Lombard, V.; Svacha, P.; Denux, O.; Augustin, S.; Henrissat, B.; McKenna, D. D. et al.: *Larvae of longhorned beetles (Coleoptera; Cerambycidae) have evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes.* Systematic Entomology 46 (4), pp. 784 - 797 (2021)

Häger, W.; Wielsch, N.; Shin, N. R.; Gebauer-Jung, S.; Pauchet, Y.; Kirsch, R.: *New players in the interaction between beetle polygalacturonases and plant polygalacturonase-inhibiting proteins: Insights from proteomics and gene expression analyses.* Frontiers in Plant Science 12, 660430 (2021)

In Preparation

Shin, N. R.; Okamura Y.; Kirsch R.; Pauchet Y.: *Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles.* Submitted to Insect Molecular Biology, Status: under reviewed

Presentation

Shin N.R. (2021). GH5_2 as almighty secret weapon turned to be essential enzymes in cerambycidbeetles; 20th IMPRS Symposium, International Max Planck Research School "The Exploration of Ecological Interactions with Molecular and Chemical Techniques" (virtual), Jena, Germany, Apr 2021

Shin N.R. (2019) Distribution of plant cell wall degrading enzymes in beetles of the family Cerambycidae at: 18th IMPRS Symposium, International Max Planck Research School, Dornburg, DE, Mar 2019

Poster presentations

Shin N.R., Shin S., Kirsch R., Svacha P., Denux O., Augustin S., Lombard V., Henrissat B., McKenna D., Pauchet Y. (2019). Transcriptome analyses provide an updated phylogeny of the family Cerambycidae and insights into the distribution of plant cell wall degrading enzymes in these beetles. Institute Symposium, Max Planck Institute for Chemical Ecology, Jena, DE (2019)

Shin N.R., Shin S., Kirsch R., Svacha P., Denux O., Augustin S., Lombard V., Henrissat B., McKenna D., Pauchet Y. (2019). Transcriptome analyses provide an updated phylogeny of the family Cerambycidae and insights into the distribution of plant cell wall degrading enzymes in these beetles. 9th Dresden Meeting Insect Phylogeny, Dresden, Germany, Sep, 2019

Shin N.R., Pauchet Y. (2019). Distribution of plant cell wall degrading enzymes in beetles of the family Cerambycidae. 18th IMPRS Symposium, Max Planck Institute for Chemical Ecology, IMPRS, Dornburg, DE

Shin N.R., Pauchet Y. (2018). Distribution of plant cell wall degrading enzymes in beetles of the family Cerambycidae. XI European Congress of Entomology, Naples, IT

Shin N.R., Pauchet Y. (2018). Diversity, evolutionary history and functional characterization of plant cell wall degrading enzymes in beetles of the family Cerambycidae. 17th IMPRS Symposium, International Max Planck Research School, Dornburg, DE

E. Declaration on honour

I hereby declare that I am familiar with the valid doctoral examination regulations of the Faculty of Biological Sciences of Friedrich Schiller University Jena. I wrote this dissertation independently with the assistance and literature cited. All contributions by people and organizations for experiments, evaluation, and preparation of the manuscripts are clearly listed by name.

I declare that I did not get any assistance from specialized consultants and a doctoral advisor, and any third parties have not received any direct or indirect benefits from me for the work connected to the doctoral thesis submitted.

The dissertation has not been previously submitted whether to the Friedrich Schiller University Jena or another university.

_____

Na Ra Shin

Jena 29.09.2022