Plant cell wall degradation in Coleoptera: Investigation of three glycoside hydrolase families implicated in cellulose and hemicellulose digestion in Phytophaga beetles

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 the Friedrich Schiller University Jena

by

Dipl. Biol. André Busch

born on June 11th, 1981 in Jena

Das Promotionsgesuch wurde eingereicht und bewilligt im: Januar 2019

Gutachter:
1. Prof. David G. Heckel (Max-Planck-Institut für chemische Ökologie)
2. Prof. Günther Theißen (FSU Jena, LS Genetik)
3. Dr. Marie-Noëlle Rosso (INRA, French National Institute for Agricultural Research)

Das Promotionskolloquium wurde abgelegt am: 24.Juli 2019

“Do or do not. There is no try.”

--Yoda, Jedi Master

Contents

1. General Introduction

1.1. The plant cell wall

Each plant cell is wrapped in a protective layer known as the plant cell wall (Fig. 1). A flexible and plastic primary cell wall allows developing cells to grow. In contrast, mature cells that have stopped growing develop a secondary cell wall by incorporating lignin, which provides further stability and resistance against outside stressors. On the molecular level, the plant cell wall consists of an array of sugar-based polysaccharides synthesized by the plant cell. The predominance of polysaccharides in the plant cell wall is directly linked to the plant's ability to perform photosynthesis, a process which permits an almost unlimited supply of sugars to be used as building blocks for a complex and resistant wall structure. In turn the plant can use more limited resources such as nitrogen or phosphate (two major plant growth factors) for growth (Pauly and Keegstra 2008). Traditionally the plant cell wall consists of three major, chemically diverse, polysaccharides: namely, cellulose, hemicelluloses and pectins. These polysaccharides represent a proportion of 20-50%, 15-35 % and 5-35% of the plant cell wall composition, respectively (Payne et al. 2015). All three polysaccharides represent the most abundant biopolymers on earth and are of major interest for the food-, textile-, paper- and biofuel industries (Pauly and Keegstra 2008). In the following chapters a closer look at each of those polysaccharides will be given.

1.1.1. Cellulose

Cellulose consists of a straight chain of β-1,4-linked glucose moieties. Individual chains emerge from the cellulose synthase complex in parallel, attached to each other by hydrogen bonding and ultimately resulting in an aggregate of (para)crystalline cellulose microfibrils (Chang 1981; Somerville 2006; Thomas et al. 2013). Due to the incorporation of lignin (which in turn replaces water), levels of cellulose crystallinity are believed to be high in the secondary plant cell wall (Pauly and Keegstra 2008).

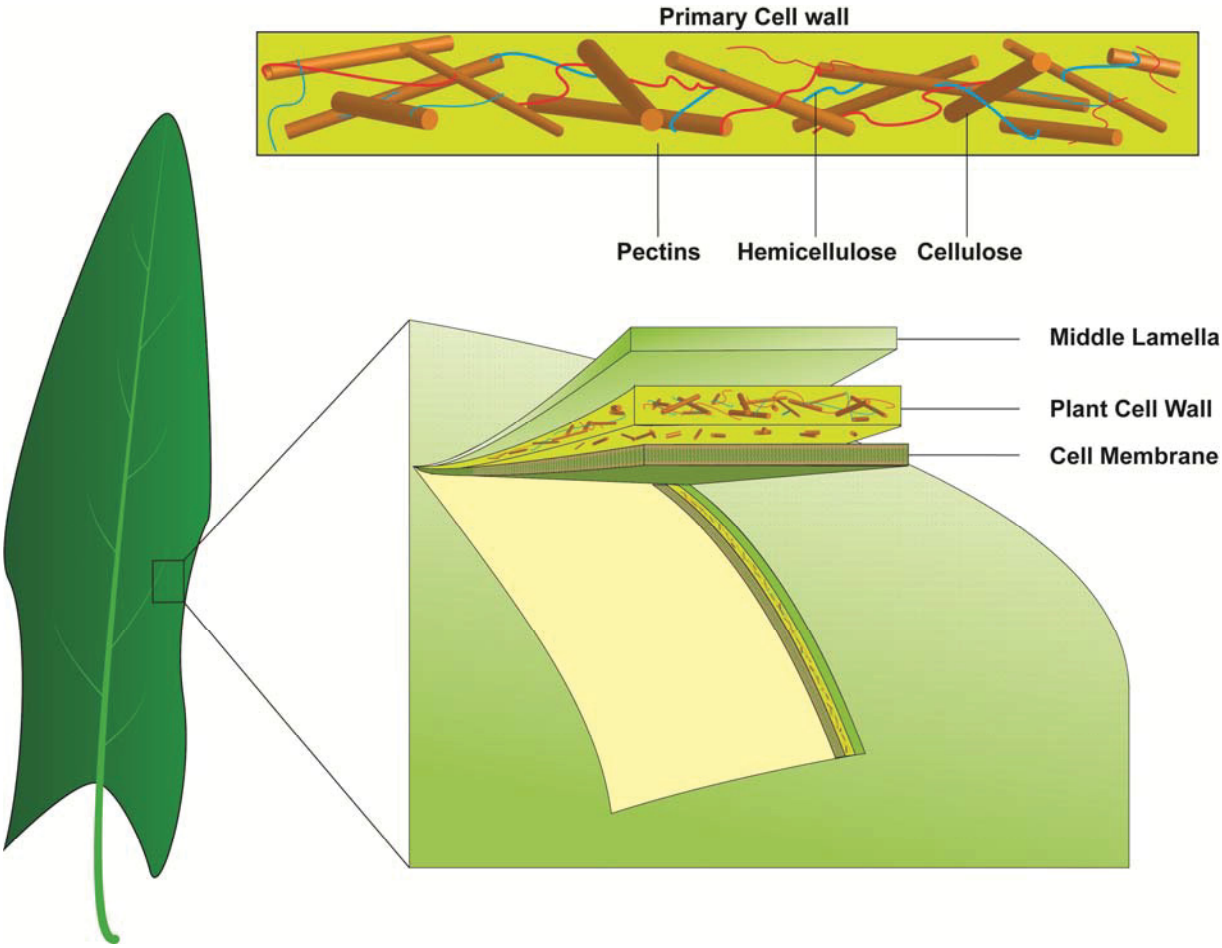


Fig. 1: Schematic of the plant outer cell region. Depicted is a partial plant cell dissected into its three surface layers: middle lamella, cell wall and cell membrane as well as the major polysaccharides of the cell wall.

Interestingly, not all cellulose parts are water-free, in particular in surface areas where cellulose microfibrils are in conformational disorder and are able to form hydrated “amorphous” areas (Fig. 2) (Beguin and Aubert 1994; Viëtor et al. 2002). Additionally, it is suspected that other polysaccharides (e.g. hemicelluloses) become entrapped within the cellulose microfibril formation, resulting in the disruption of the cellulose crystal structure (Cosgrove 2005). Within the primary cell wall, it is hypothesized (in addition to the above-mentioned facts) that “sloppy packing” and the random distribution of cellulose microfibrils during cell growth contribute to the formation of amorphous zones (Cosgrove 2014; Montanari et al. 2005). However, despite efforts to resolve the distribution of amorphous cellulose in relation to its crystalline state in the primary and secondary plant cell wall, much remains unknown (Ruel et al. 2012).

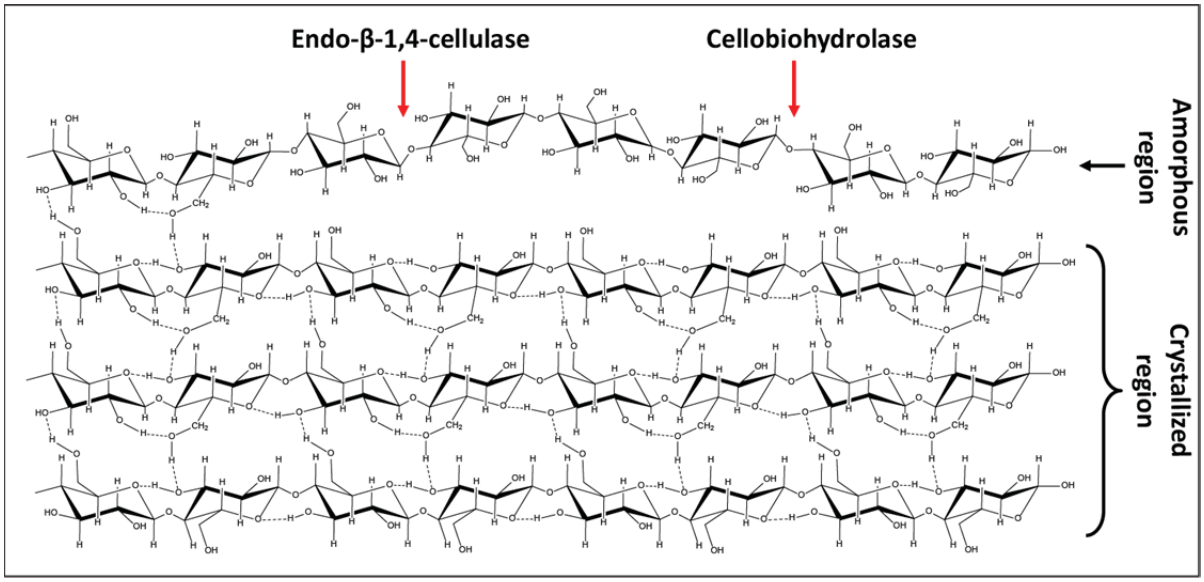


Fig. 2 Chemical structure and degradation of crystalline and amorphous cellulose. Amorphous cellulose is characterized by water solubility and is believed to be primarily located at surface areas of cellulose microfibrils. Endo-β-1,4,-glucanases (EC 3.2.1.4) are able to cleave cellulose in its amorphous state by releasing random-sized blocks of cellulose. In contrast, crystalline regions feature hydrogen bonds between individual cellulose chains that are made by excluding water and forming a quasi-crystalline structure. Cellobiohydrolases (EC 3.2.1.91) are able to cleave at the ends of the crystalline and amorphous regions. Red arrows indicate potential cleaving sites for the respective cellulases.

In plants, cellulose can represent between 20 to 50% of the polysaccharides and is therefore believed to be the most abundant biopolymer on earth, making it a major research focus of the biofuel industry, which aims to convert cellulosic biomass into fuel-supplemented ethanol.

1.1.2. Hemicellulose

Hemicelluloses are a diverse set of heteropolymers, being structurally and physicochemically unrelated to each other. Based on Scheller and Ulvskov (2010), hemicelluloses mainly consist of the heteropolysaccharides xylan and xyloglucan, as well as mannans (glucomannan, galactomannan and galactoglucomannan). In contrast to cellulose, hemicelluloses are prevented from forming microfibrils due to their side-chain residues; these allow hydration and, conversely, inhibit crystallization (Gilbert 2010; Thomas et al. 2013). Within the plant cell wall, hemicelluloses are

thought to be mechanically linked to cellulose to further stabilize the structure (Cosgrove 2014; Scheller and Ulvskov 2010).

Xylan consists of a backbone of xylose units and is commonly modified by substitutions of glucuronosyl residues (glucuronoxylan) and/or arabinose residues (arabino- and glucuronoarabinoxylan) (Fig. 3). The former can be found as major non-cellulosic polysaccharide in the secondary plant cell wall of dicots, whereas the latter dominates the primary plant cell wall of certain monocots (Scheller and Ulvskov 2010). Other side-chain substitutions of xylan include, for example, galactose, methylglucuronosyl residues or rhamnose (Javier et al. 2007).

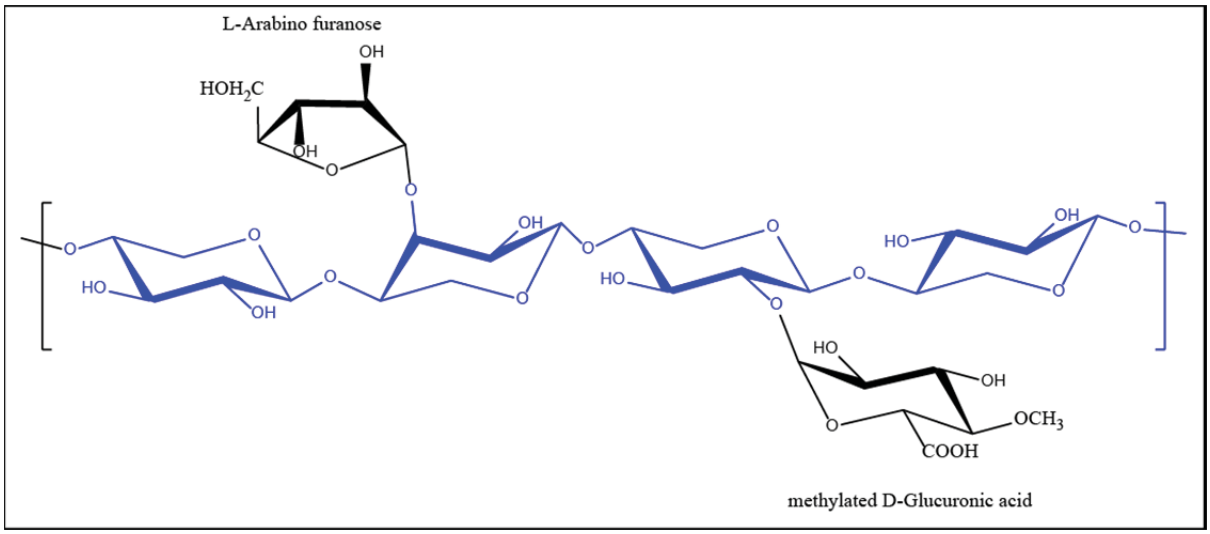


Fig. 3 Chemical structure of xylan. Xylan can be highly modified by side-chain residues. Two potential residues are depicted here. Blue represents xylose units connected in a β-1,4-manner, which forms the backbone xylan.

Xyloglucan is represented in all land plants and is assumed to be the major hemicellulose in the primary plant cell wall. Its general structure consists of beta-1,4-linked glucose residues (similar to cellulose) which are unequally substituted with xylose units (Fig. 4). In turn xylose units can be further substituted with D-galactose, L-fucose and D-arabinose among others. The degree of substitution influences the solubility of xyloglucan, which varies greatly between plant species.

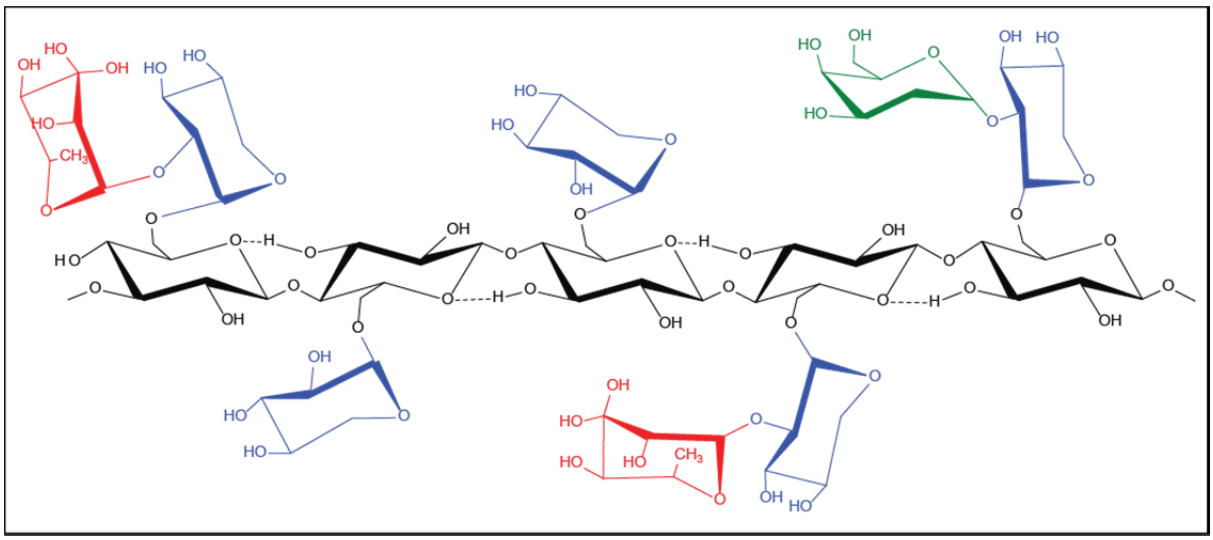


Fig. 4 Chemical structure of xyloglucan. As a close molecular relative to cellulose, its backbone consists of D-glucose residues linked by a β-1,4 glycosidic bond. Depicted in black are D-glucose residues; in blue, D-xylose; in red, L-fucose; and in green, D-galactose.

The backbone of mannans consists of mannose residues (pure mannan), which in turn can be substituted with galactose residues (galactomannan). In contrast, glucomannan has a backbone of randomly distributed glucose and mannose residues which can be occasionally substituted with galactose (galactoglucomannan) (Scheller and Ulvskov 2010). Mannans often occur as storage polysaccharides in seeds, bulbs, tubers and roots (Buckeridge et al. 2000; Meier and Reid 1982), but they are also readily incorporated in the plant cell wall as structural components (Cosgrove 2005; Scheller and Ulvskov 2010).

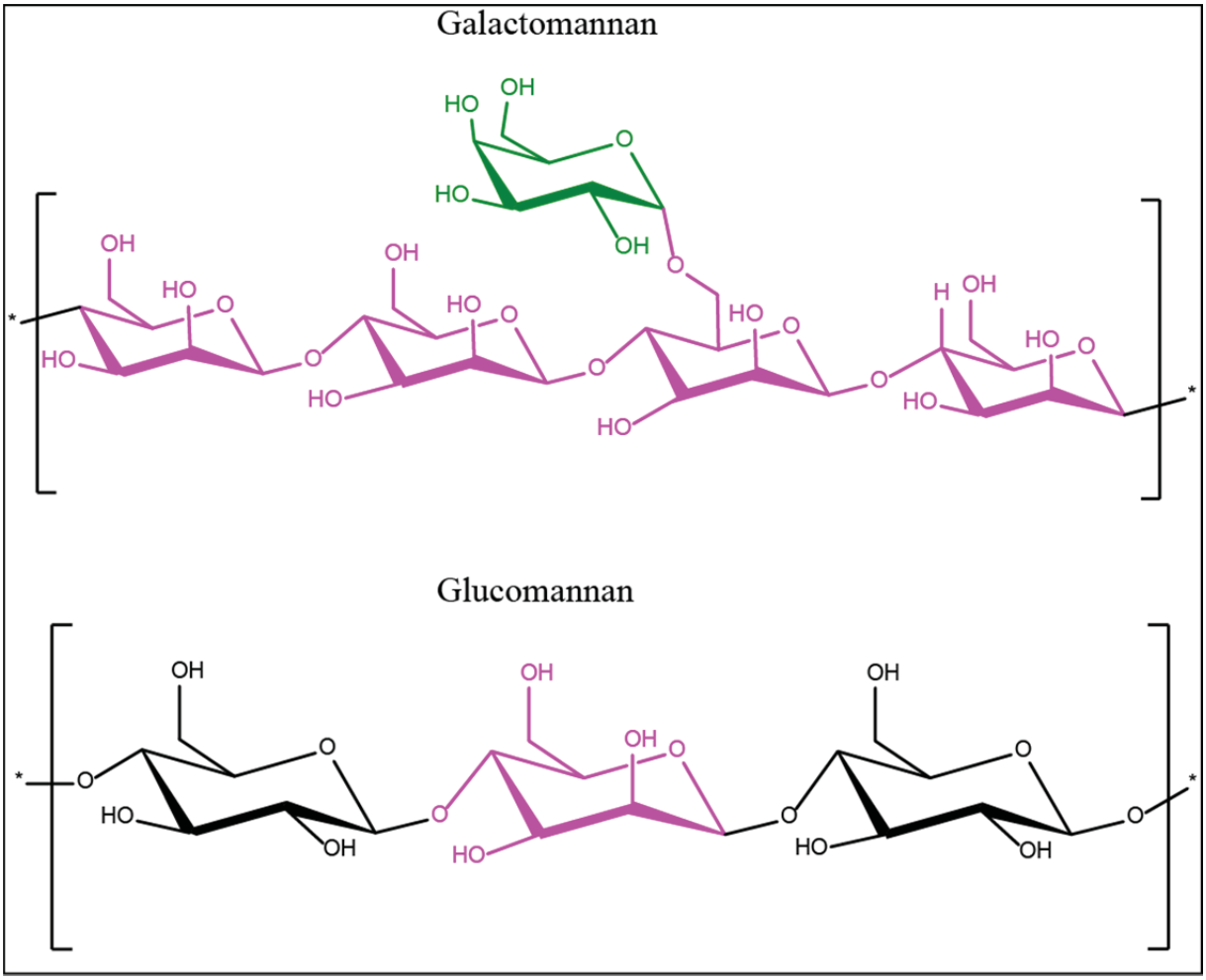


Fig. 5 Chemical structure of galacto- and glucomannan. Galactomannan consists of a backbone of D-mannose residues linked by β-1,4 glycosidic bonds that are decorated with D-galactose units. In contrast, glucomannan is a straight-chain polymer consisting of unevenly distributed D-glucose and D-mannose residues. Occasional branching with D-galactose is common (galactoglucomannan). Pink = D-mannose residues, green = D-galactose, black = D-glucose.

1.1.3. Pectins

Pectins are a group of heterogeneous polysaccharides encompassing several complex polymers. The least complex is homogalacturonan, comprising of a straight chain of galacturonic acid moieties which can be substituted by xylose residues (xylogalactorunan). Rhamnogalacturonan I is composed of altering rhamnose and galacturonic acid residues with eventual branching, whereas rhamnogalacturonan II is a highly complex aggregate containing up to 11 different sugar residues. Pectins

embed both cellulose and hemicellulose, and are believed to increase the flexibility of the plant cell wall (Cosgrove 1997; Cosgrove 2005; Keegstra 2010).

1.2. Cellulolytic systems in animals

Any organism able to degrade the plant cell wall gains access to many of plant-cell-wall-derived sugars. Besides obtaining a substantial boost of energy, the degradation of the plant cell wall allows for the exploitation of the cell content, providing additional resources of nitrogen which are ultimately required for growth (Kainulainen et al. 1996; Rossi et al. 1996). Ironically, plant cell wall polysaccharides form a resilient network which is resistant to enzymatic breakdown (in contrast to the highly similar, yet easily digestible storage polysaccharides, such as amylose or starch), only allowing highly specified organisms to degrade the plant cell wall (Bayer et al. 1998). Classically, the degradation of plant cell wall material by the use of so-called plant-cell-wall-degrading enzymes (PCWDEs) was attributed only to the domain of microbes. PCWDEs were found in plant pathogenic bacteria and fungi, saprotrophs and in symbionts in the gut of herbivorous animals (Breznak and Brune 1994; Chambost J.P. 1987; Rincon et al. 2001; Schulein 1997).

Conversely, it was commonly accepted that the success of herbivorous animals was ensured by the presence of mutualistic plant-cell-wall-degrading microbes in their guts. The first cracks in the belief that the cellulolytic system was solely dependent on symbionts did not arise until 1978, when a symbiont-free marine isopod was found to be still able to degrade cellulose (Boyle and Mitchell 1978; King et al. 2010). Despite this early discovery it took scientists another 20 years to sequence and characterize the first endogenous cellulase in a wood-feeding termite (Watanabe et al. 1998), and shortly after in nematodes (Smant et al. 1998). That initial discovery was followed by the identification of endogenous cellulases in other Metazoa, most prominently Arthropoda (Calderon-Cortes et al. 2010; Kim et al. 2008; McKenna et al. 2016; Mei et al. 2016; Pauchet et al. 2014a; Shelomi et al. 2016; Willis et al. 2011), as well as in Nematoda (Kikuchi et al. 2004; Palomares-Rius et al. 2014) and Mollusca (Guo et al. 2008; Sakamoto and Toyohara 2009; Tsuji et al. 2013). Less frequently, endogenous cellulases have also been identified in the phyla Echinodermata (Nishida et al. 2007),

Chordata (Davison and Blaxter 2005; Dehal et al. 2002), Annelida (Arimori et al. 2013), Rotifera (Szydlowski et al. 2015) and Tardigrada (Davison and Blaxter 2005).

Based on their activity, cellulases are separated into endo-β-1,4-glucanases (EC 3.2.1.4) and cellobiohydrolases (exoglucanases; EC 3.2.1.91). The former class of enzymes cleaves amorphous cellulose randomly, releasing fragments of oligopolysaccharides of varying sizes (Fig. 2). Cellobiohydrolases cleave cellulose from its reducing or non-reducing end by releasing dimers and occasionally trimers (Fischer et al. 2013; Takahashi et al. 2010). Cellobiohydrolases can degrade both amorphous and crystalline structures of the cellulosic network (Liu et al. 2011; Shen et al. 1995).

According to their amino acid structure, cellulases can be classified into distinct families of glycoside hydrolases (GHs). It has been found that cellulases (i.e. endo- and exo-acting cellulases) do not belong to a single GH family but appear in diverse sets including the families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 74 and 124 (www.cazy.org) (Lombard et al. 2014). Notably, out of those 14 GH families only seven have been identified to be encoded by animals: GH5, GH6, GH7, GH9, GH10, GH45 and GH48. All of these GH families consist of a catalytic dyad harboring a glutamate and/or an aspartate. Both amino acids can perform either as a catalytic acid or as a base, e.g. in GH45s both catalytic residues are aspartates, whereas in GH5s both residues are glutamates. Upon hydrolysis the emerging C1 hydroxyl group may be retained in its original configuration or it may be inverted, facing the opposite site of its original configuration. Therefore, the hydrolytic mechanism is referred to as either “retaining” (e.g. GH5s, GH7s) or “inverting” (e.g. GH9s, GH45s) mode of action (Fig. 6).

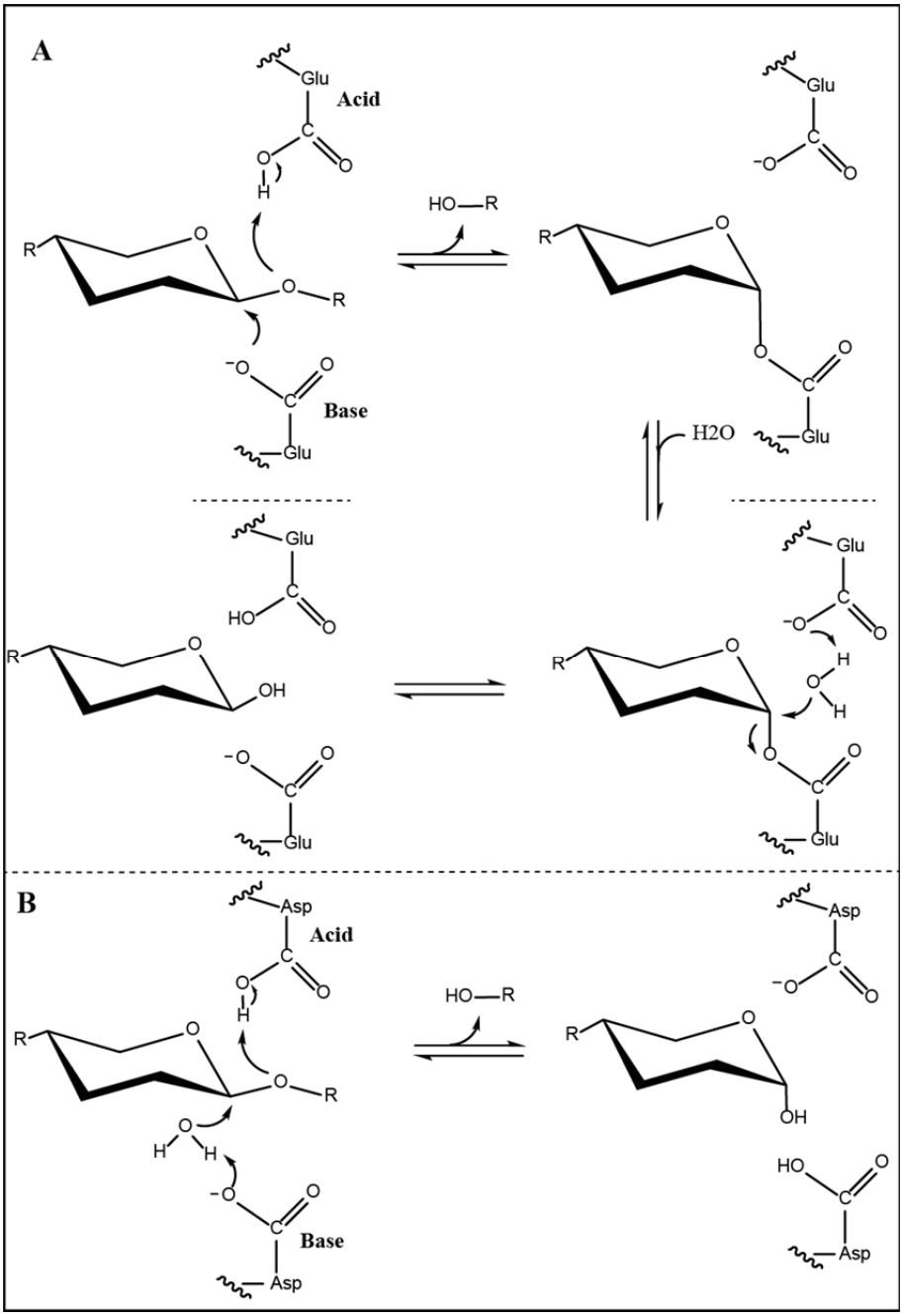


Fig. 6 Catalytic mechanism of glycoside hydrolases comparing retaining and inverting modes of action. A) Depicts the retaining mechanism of glycoside hydrolases with two glutamates as catalytic core residues (e.g. in GH5 family members). After substrate cleavage, the C1-hydroxyl-group of the glucose residue is retained in its β-configuration. B) Depicts the inverting mechanism of glycoside hydrolases with two aspartates as catalytic core residues (e.g. in GH45 family members). After substrate cleavage, the C1-hydroxyl-group of the glucose residues has inverted from β- to α-configuration. Adapted from Davies and Henrissat (1995).

In microorganisms, cellulases are frequently coupled to non-catalytic carbohydrate binding domains (CBMs) and form, together with several scaffolding proteins, multi-enzyme complexes (Matte et al. 2009). These complexes can be found as small protein aggregates consisting of a single cellulolytic enzyme attached to a single CBM but can also appear as extremely large protein aggregates consisting of several CBMs and catalytic enzymes (so-called cellulosomes) (Gaudin et al. 2000; Sukharnikov et al. 2012; Watanabe and Tokuda 2010). Within those complexes, CBMs are believed to facilitate substrate cleavage, by i) increasing the enzyme concentration on the polysaccharide surface, ii) targeting specific substrate regions and iii) disrupting insoluble polysaccharide structures such as crystalline cellulose (Bolam et al. 1998; Boraston et al. 2004; Guillen et al. 2010). Removal of the CBM may lead to reduction or complete loss of enzymatic activity on insoluble substrates but has less effect on the activity on soluble substrates. Interestingly, and in contrast to their microbial counterparts, cellulases in insects have not generally been found to contain a CBM or multi-enzyme complex. A potential loss of the auxiliary domains of animal cellulases suggests that either they have developed other strategies to efficiently degrade crystalline cellulose or they do not degrade it at all.

1.2.1. Cellulases of the GH5 subfamily 2 (GH5_2)

The GH5 family is one of the largest and most diverse GHs, with two glutamates as catalytic residues responsible for retaining hydrolysis. GH5s are classified into 51 subfamilies (Aspeborg et al. 2012) and to date only GH5_2 has been reported in animals as able to degrade cellulose. With the exception of having evolved two novel enzymatic functions in *Anoplophora glabripennis* -- namely, xylanase and xyloglucanase activity (McKenna et al. 2016) -- GH5_2 cellulolytic function appears to be conserved. Notably, their distribution in animals is patchy and is thought to be limited to the Phyla Nematoda (Karim et al. 2009; Ledger et al. 2006; Smant et al. 1998) and Arthropoda (McKenna et al. 2016; Pauchet et al. 2014a; Sugimura et al. 2003). Interestingly, within Arthropoda GH5_2 is so far restricted to xylophagous beetles of the Cerambycidae (Pauchet et al. 2010). The patchy distribution of genes encoding proteins of GH5_2 therefore raised question about their ancestral origin in animals. Intriguingly, research on nematode-derived GH5_2 has provided evidence

that this gene subfamily is not ancestral but has evolved independently (Danchin et al. 2010). It has been suggested that these genes were likely acquired by horizontal gene transfer (HGT) from a bacterial donor. However, whether GH5_2 evolved the same way in Cerambycidae remains unresolved (Danchin et al. 2010).

1.2.2. Multifunctional cellulases of the GH9 family

GH9 was the first glycoside hydrolase family to be identified as an endogenous cellulase in animals (Watanabe et al. 1998). Since then, GH9s have received wide attention and have been identified in a broad range of animal phyla, including Arthropoda, Annelida, Echinodermata, Chordata and Mollusca. The majority of GH9s to date have been identified in arthropods, including numerous GH9 encounters in Insecta (Busch et al. 2018b; Kim et al. 2008; Watanabe et al. 1998; Willis et al. 2011) and also in Crustacea (Bui and Lee 2015; Colbourne et al. 2011; Davison and Blaxter 2005). Similarly, GH9-encoding genes have been uncovered in Mollusca (Li et al. 2009; Suzuki et al. 2003; Wang et al. 2017; Zhang et al. 2012) and Chordata (Davison and Blaxter 2005; Dehal et al. 2002) but remained rare in Annelida (Arimori et al. 2013), Echinodermata (Nishida et al. 2007) or Rotifera (Szydlowski et al. 2015). Discoveries in the latter three phyla have been scarce, likely because of a lack of general scientific attention throughout the last decades, yet the phyla remain promising locations for future GH9 identification. Additionally, GH9s are found in plants (Libertini et al. 2004), bacteria (Gaudin et al. 2000) and fungi (Nagy et al. 2016). The broad distribution of GH9 in all these animal lineages, as well as in plants and microbes, suggests that GH9s may have been present in an ancestor of bacteria and eukaryotes (Davison and Blaxter 2005).

GH9s are characterized by their inverting catalytic mechanism with an aspartate as catalytic base and a glutamate as the corresponding catalytic acid. This family is known to act as an endocellulase and/or a cellobiohydrolase, and was observed additionally to act as a β-glucosidase (www.CAZy.org). Intriguingly, and although GH9s have been characterized mostly as cellulases, other activities have evolved most notably against xyloglucan and xylan (Lombard et al. 2014; Pauly et al. 2013; Shelomi et al. 2016).

1.2.3. GH6, GH7 and GH10 cellulases

Much rarer in animals than the previously mentioned GH families are GH6, GH7 and GH10 family members. Interestingly, genes encoding GH6 proteins were not identified as single enzymes but as part of a cellulose synthase complex of marine sea squirts (Matthysse et al. 2004; Nakashima et al. 2004). Although both research groups performed no further characterization experiments, they independently found crucial mutations in the catalytic residues which led to the conclusion that GH6s derived from sea squirts have likely lost their cellulolytic activity. Ultimately, it remains to be shown whether more animal-derived GH6 members occur and whether these have truly lost their catalytic function on cellulose.

GH7-encoding genes were recently identified in a broad range of marine Crustacea, including the basal classes Branchiopoda and Copepoda (Chang and Lai 2018; Kern et al. 2013; King et al. 2010; Kobayashi et al. 2018) as well as in a slime mold (Amoebozoa) (Kunii et al. 2014). Currently, the only functionally characterized GH7 in animals was described in the crustacean isopod *Limnoria quadripunctata* (Kern et al. 2013) and the slime mold *Dictyostelium discoideum*. In both cases cellobiohydrolase activity was detected, which is in accordance with activities known from microbial GH7s. Although to date GH7-encoding genes from animals have been found mainly in Crustacea, phylogenetic evidence suggests a vertical transition from a last common ancestor of the Metazoa (King et al. 2010). If that is true, future research should unravel endogenous GH7s in a broad range of metazoans as well as microbes.

Finally, several members of the GH10 family have been described in animals. In the Mollusca *Ampullaria crossean* a GH10 was characterized (EGX) with apparent activity against cellulose as well as xylan (Wang et al. 2003). Two GH10 genes (Pc1 and Pc3), identified in the golden apple snail *Pomacea canaliculata,* were orthologous to EGX of *A. crossean* (Imjongjirak et al. 2008). Based on their orthologous relationship, the authors claimed cellulolytic activity in the corresponding proteins. However, this information should be viewed cautiously as the authors did not directly functionally characterize these GH10s. Other cases of GH10s in animals emerged,

such as in the weevil *Hypothenemus hampei*, in which the GH10 protein was characterized as an endo-β-1,4-xylanase (Padilla-Hurtado et al. 2012), or the bdeloid rotifer *Adineta ricciae* (Szydlowski et al. 2015). However, in the latter case no functional characterization was performed. Curiously, cellulase activity for this particular family is rare as GH10s are generally associated with xylanase activity. It is therefore likely that the above-observed cellulolytic activity has evolved recently based on the chemical similarities between cellulose and xylan. Nonetheless, future discoveries will likely shed light on the elusive enzymatic nature of GH10 putative cellulases in animals.

1.2.4. GH45 cellulases

Members of the GH45 family, which were found to be omnipresent in cellulolytic microbes (Gilbert et al. 1990; Saloheimo et al. 1994; Sheppard et al. 1994), contained two aspartates as catalytic residues and an inverting enzymatic mechanism (Davies et al. 1995). A GH45 gene endogenous to an animal was first described in the leaf beetle *Phaedon cochleariae* in 1999 (Girard and Jouanin). However, no direct correlation between the cellulolytic activity observed in the gut of *P. cochleariae* and the respective GH45 transcript was established. Thus, the first endogenous GH45 functionally characterized in an insect was encoded by the longhorned beetle *Apriona germari* (Cerambycidae: Lamiinae). Based on its ability to degrade amorphous cellulose, this GH45 was classified as an endo-active cellulase (Lee et al. 2004). Simultaneously, endogenous GH45 endo-β-1,4-cellulases were characterized in the pine wood nematode *Bursaphelenchus xylophilus* (Kikuchi et al. 2004). In the following years, the identification and characterization of GH45s remained limited to cerambycid beetles including *Oncideres albomarginata chamela*, *Anoplophora glabripennis* and *Apriona japonica* (Calderon-Cortes et al. 2010; McKenna et al. 2016; Pauchet et al. 2014a).

With the advance of novel sequencing strategies, GH45 transcripts also emerged in other insects. The majority of GH45s were identified in various phytophagous beetles from the family Chrysomelidae and from the superfamily Curculionoidea (Pauchet et al. 2010). However, to date only one GH45 cellulase has been functionally

characterized from a beetle of the Chrysomelidae, namely, the Western corn rootworm *Diabrotica virgifera virgifera* (Valencia et al. 2013); none has been characterized in the Curculionoidea. Apart from beetles, GH45s have been characterized in arthropods other than insects, i.e. Collembola (Song et al. 2017) as well as in several Mollusca (Rahman et al. 2014; Sakamoto and Toyohara 2009; Xu et al. 2001) and Nematoda (Wang et al. 2018). All characterized GH45s are classified as endo-acting cellulases.

Recently, GH45s have been identified in a Crustacea as well as in a species of bdeloid rotifer. However, functional data has only been provided for *A. ricciae* showing endo-β-1,4-glucanase activity (Chang and Lai 2018; Szydlowski et al. 2015). Interestingly, phylogenetic analyses of GH45s derived from nematodes and molluscs suggested that their origin was not ancestral but likely inherited from a foreign source by HGT from a fungal donor (Palomares-Rius et al. 2014; Sakamoto and Toyohara 2009). Based on this research and the patchy distribution of GH45s in arthropods (and the Metazoa in general), a similar hypothesis was assumed for beetle-derived GH45s. Yet, the first phylogenetic analysis of beetles-derived GH45s remained inconclusive due to a lack of sufficient sequence data (Calderon-Cortes et al. 2010). In a second more comprehensive approach, similar results were obtained (Eyun et al. 2014). Although the authors of the latter study increased the number of GH45 sequences from the former study, GH45 variety on species level remained low, preventing a clear evolutionary history of the GH45 gene family from being developed. To date the evolution of the GH45 family in beetles (and arthropods in general) remains unresolved. In a previous survey, we discovered several transcripts encoding putative cellulases of the GH45 family in a diverse set of Phytophaga beetles' (Curculionoidea and Chrysomeloidea; (Marvaldi et al. 2009)); seven of these were chosen to be investigated in more detail. Six of those beetles belong to the Chrysomelidae (leaf beetles) and one to the Curculionidae (weevils). GH45 transcripts encoded by beetles of interest are depicted in Figure 6.

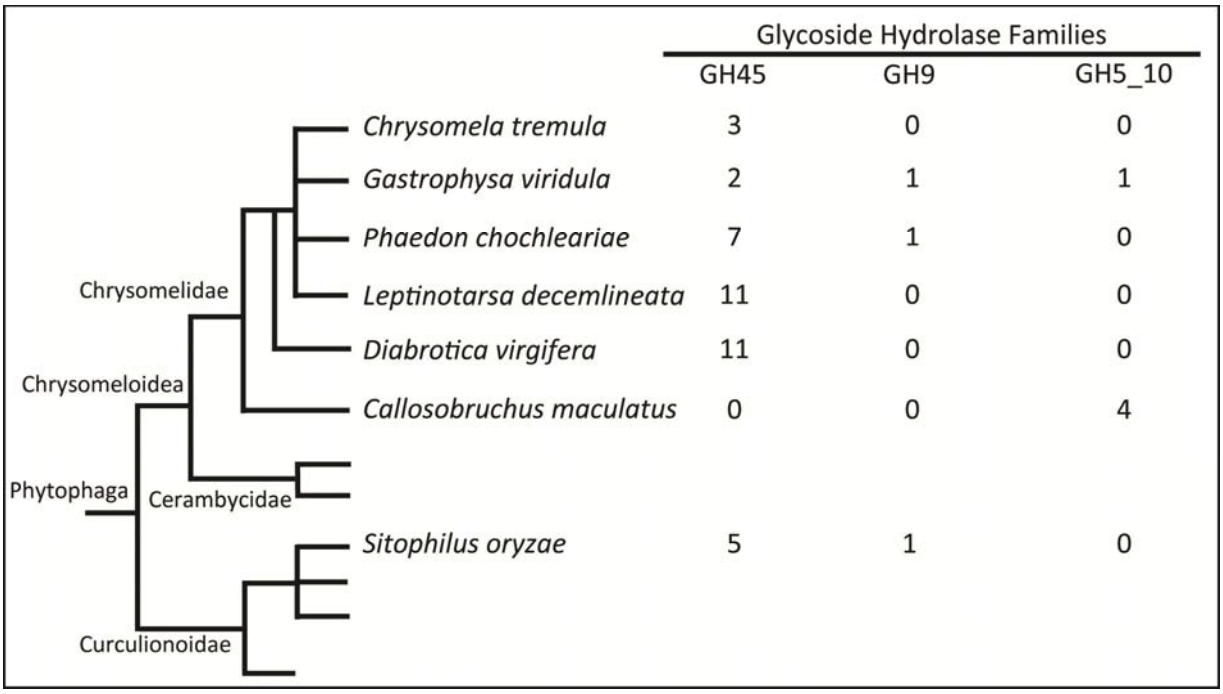


Fig. 6 Target beetles and the number of their transcripts that encode the putative PCWDEs investigated in this thesis. Adapted from Pauchet et al. (2010).

More precisely our research focus is *Chrysomela tremula* (Chrysomelidae: Chrysomelinae), which mainly feeds on poplar; *Gastrophysa viridula* (Chrysomelidae: Chrysomelinae), which feeds on *Rumex spp*. plants; *Phaedon cochleariae* (Chrysomelidae: Chrysomelinae), which feeds on brassicaceous plants; *Leptinotarsa decemlineata* (Chrysomelidae: Chrysomelinae), which is notorious for feeding on potato plants and on other solanaceous species; *Diabrotica virgifera virgifera* (Chrysomelidae: Galerucinae), a notorious pest on maize; *Callosobruchus maculatus* (Chrysomelidae: Bruchinae), a pest species on legumes; and *Sitophilus oryzae* (Curculionidae: Dryophthorinae), a pest on rice and other stored grains.

1.2.5. GH48 cellulases

Genes encoding putative cellulases of the GH48 family have been identified in Phytophaga beetles and represent the only examples to date of this GH family found in animals (Pauchet et al. 2010). Their catalytic core consists of a glutamate as proton donor (Parsiegla et al. 1998) and an aspartate as proton acceptor (Kostylev and Wilson 2011). Based on functional characterization experiments in microbes, GH48s have been classified as cellobiohydrolases (Irwin et al. 2000; Sanchez et al. 2003) or

endoglucanases (Reverbel-Leroy et al. 1997). To date, only two attempts have been made to functionally characterize GH48s in Coleoptera. First, a GH48 from *Otiorhynchus sulcatus* (Coleoptera: Curculionidae), resulting in inconclusive cellobiohydrolase activity (Edwards 2002), and second, a GH48 from *Gastrophysa atrocyanea* (Coleoptera: Chrysomelidae), which was surprisingly but not unambiguously characterized as a chitinase (Fujita et al. 2006). Based on these results, the function of GH48 in insects needs to be more thoroughly addressed to determine whether they are cellobiohydrolases, endoglucanases or chitinases.

1.3. Hemicellulolytic systems in animals

As stated above, hemicelluloses consist mainly of xyloglucan, xylan and mannans (Scheller and Ulvskov 2010). Conversely, enzymes that are able to degrade those substrates are endo-β-1,4-xyloglucanases (EC 3.2.1.151), endo-β-1,4-xylanases (EC 3.2.1.8) and endo-β-1,4-mannanases (EC. 3.2.1.78). According to the CAZy database (www.cazy.org), these enzyme classes are distributed in a wide range of GH families within the tree of life (Lombard et al. 2014). However, endo-β-1,4-xyloglucanase activity in animals has been described only once for a functionally highly diverse group of GH9s, in stick insects (Phasmatodea) (Shelomi et al. 2016), and another time for a GH5_2, in the longhorned beetle (*A. glabripennis)* (McKenna et al. 2016). Likewise, endo-β-1,4-xylanase activity in animals is rare and has been described only for two GH11s encoded by the leaf beetle *Phaedon cochleariae* (Pauchet and Heckel 2013), for GH9s in Phasmatodea (Shelomi et al. 2016), for GH5_2 in Cerambycidae (McKenna et al. 2016) and for a single GH10 in the coffee berry borer *H. hampei* (Padilla-Hurtado et al. 2012). Finally, endo-β-1,4-mannanases have been described in several of the Metazoa but are limited to two subfamilies of GH5, namely subfamily 8 (GH5_8) and subfamily 10 (GH5_10). To date, the single GH5_8 characterized in *H. hampei* displayed endo-β-1,4-mannanase activity (Acuna et al. 2012b). All additional endo-β-1,4-mannanases were members of GH5_10, which will be further described in the following section.

1.3.1. Endo-β-1,4-mannanases of GH5 subfamily 10

Although recently discovered in several Crustacea (Gan et al. 2018), GH5_10 proteins have been identified and functionally characterized in only a handful of other species, including the bivalve *Mytilus edulis* (Larsson et al. 2006; Xu et al. 2002), the gastropods *Haliotis discus hannai* and *Aplysia kurodai* (Ootsuka et al. 2006; Zahura et al. 2011), the Crustacea *Daphnia pulex* (King et al. 2010), the annelid *Eisenia fetida* (Ueda et al. 2018) and the Collembola *Cryptopygus antarcticus* (Song et al. 2008). All of these GH5_10s were characterized as endo-β-1,4-mannanases. Based on their patchy distribution, the complex evolutionary history of the GH5_10 family is still under investigation (Tan et al. 2016). According to transcriptome data (Pauchet et al. 2010), GH5_10 putative mannanases are encoded by two beetles of the Chrysomelidae, the green dock beetle (*G. viridula)* and the bean beetle (*C. maculatus)*. This discovery represents a rare finding in beetles, as the only other mannan-degrading enzyme found in a beetle species belonged to a different subfamily of the GH5 (Acuna et al. 2012b). Conversely, these findings suggest that mannan degradation has evolved at least twice in Phytophaga beetles. Additionally, the patchy distribution of the GH5_10 members in insects and animals, in general, suggests that they were likely not transferred vertically from an ancestral origin but may have evolved independently during beetle evolution.

1.4. Aims of the study and research questions addressed in this thesis

The overall aim of this thesis is to investigate the ability of Phytophaga beetles to digest plant cell wall material and to determine the evolution of genes involved in that process. The following questions are addressed: How many genes were present in the LCA of the Phytophaga clade of beetles? Have beetle-derived PCWDEs kept their ancestral function or have they evolved novel functions? What is their evolutionary history? Is the origin of those genes ancestral or have they been acquired horizontally?

In **Manuscript one, I** focus on genes encoding GH5_10 putative mannanases identified in two beetles of the Chrysomelidae with different feeding habits. I investigate the ability of GH5_10s to degrade plant-cell-wall-derived polysaccharides, and their physiological importance as well as their evolutionary origin.

In **Manuscript two**, *Gastrophysa viridula* is the model organism through which I investigate three genes encoding glycoside hydrolases belonging to GH9 and GH45, both of which are known from microbes and Metazoa to act as cellulases. Here, I detail their enzymatic function, biochemical properties and biological relevance.

In **Manuscript three,** I investigate the evolution of GH45s in Phytophaga beetles and extend this investigation to other GH45-encoding animal Phyla. Additionally, beetle-derived GH45s are extensively characterized and the functional data obtained put into context through phylogenetic analysis.

This thesis greatly increases our understanding of PCWDEs encoded by Phytophaga beetles. We show that GH45s as well as GH5_10s have evolved to acquire novel substrate specificities that allow those beetles to degrade several plant cell wall polysaccharides by expressing only two GH families. We also shed light on the especially intricate evolution of GH45s, a family which has evolved several times throughout the history of Arthropods. Our data suggest that GH45s were acquired from a fungal source, likely through a HGT. On a broad perspective, our functional experiments provide evidence that beetles of the Phytophaga may represent a source

of novel enzymes useable in processes such as paper making, beer brewing and biofuel production.

2. Overview of Manuscripts

Manuscript I

Evolution and functional characterization of CAZymes belonging to subfamily 10 of glycoside hydrolase family 5 (GH5_10) in two species of phytophagous beetles

André Busch[a], Grit Kunert[b], David G. Heckel[a] ,Yannick Pauchet[a]

Department of [a]Entomology and [b]Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

Published in *PLoS One, Volume 12, Issue 8, August 2017, e0184305*

In **Manuscript I**, we investigated putative mannanases belonging to members of glycoside hydrolase family 5 subfamily 10 (GH5_10) in *Gastrophysa viridula* (Chrysomelidae, Chrysomelinae) and *Callosobruchus maculatus* (Chrysomelidae, Bruchinae). We focused on GH5_10 functional characterization, biological relevance and evolutionary origin. We were able to confirm GH5_10 activity against mannan-derived polysaccharides but also found novel activities against xylan and cellulose. GH5_10 gene silencing using RNAi in *G.viridula* was successful resulting in a reduced gene as well as corresponding protein level. However, we were unable to detect any changes in phenotype suggesting a complex regulation of GH5_10 in the beetle. Phylogenetic analysis revealed a paraphyletic relationship of insect-derived GH5_10 suggesting that those genes were acquired at least twice during insect evolution likely through horizontal gene transfer.

André Busch and Yannick Pauchet planned the experimental framework, performed the experiments, analyzed the data, prepared figures and wrote the manuscript. Grit Kunert performed statistical analysis and helped discussing the data. David G. Heckel participated in the discussion of the results. All authors revised the manuscript.

Manuscript II

Cellulose degradation in Gastrophysa viridula (Coleoptera: Chrysomelidae): functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity

André Busch[a], Grit Kunert[b], Natalie Wielsch[c], Yannick Pauchet[a]

Department of [a]Entomology and [b]Biochemistry and [c]Research Group Mass Spectrometry, Max Planck Institute for Chemical Ecology, Jena, Germany

Published in *Insect Molecular Biology, Volume 27, Issue 5, May 2018, Pages 633-650*

In **Manuscript II**, we investigated putative cellulases belonging to members of glycoside hydrolase family 45 (GH45) and 9 (GH9) in *Gastrophysa viridula* (Chrysomelidae, Chrysomelinae). The major aim was to functionally characterize these proteins and investigate their biological function. Functional expression revealed that one GH45 has kept its cellulolytic function whereas the other has evolved to degrade xyloglucan. No activity was detected for the GH9. Gene silencing of both GH45s using RNAi resulted in a successful gene knockdown but no phenotypic effects were observed. Zymography experiments identified several other cellulases in the gut of *G. viridula* explaining the missing phenotype. Yet, the same experiment did not detect any additional xyloglucanases indicating a complex regulation and compensation of the silenced GH45 by other digestive enzymes.

André Busch and Yannick Pauchet planned the experimental set up. André Busch performed the experiments, analyzed the data with the help of Yannick Pauchet, prepared the figures and wrote the manuscript. Grit Kunert performed statistical analysis and helped discussing the data. Natalie Wielsch performed LC-MS/MS analysis. All authors revised the manuscript.

Manuscript III

Functional analyses of the horizontally acquired Phytophaga glycoside hydrolase family 45 (GH45) proteins reveal distinct functional characteristics

André Busch[1], Etienne G.J. Danchin[2] and Yannick Pauchet[1]*

[1]Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany and
[2]INRA, Université Côte d'Azur, CNRS, ISA, Sophia Antipolis, France.

Manuscript in preparation and to be submitted to *Molecular Biology and Evolution*

In **Manuscript III**, we performed a wide spread analysis of beetle-derived GH45s encompassing four beetles of the Chrysomelidae and one beetle of the Curculionidae. We performed functional characterization experiments and identified active GH45 cellulases in each investigated species. Additionally, some GH45s have lost their ability to degrade cellulose but have evolved to degrade xyloglucan instead. Substitutions from aspartate to glutamate of catalytically important amino acid were likely responsible for the substrate shift. Our phylogenetic analysis revealed that beetle GH45s are not monophyletic with other Arthropoda GH45s but were most closely related to fungal ones suggesting a horizontal gene transfer from fungi to the last common ancestor of Phytophaga beetles.

André Busch and Yannick Pauchet planned the experimental set up. André Busch performed the experiments, analyzed the data with the help of Yannick Pauchet, prepared the figures and wrote the manuscript. Yannick Pauchet performed phylogenetic analysis on beetle-derived GH45s. Etienne Danchin performed phylogenetic analysis on animal-derived GH45s and helped discussing the results. All authors revised the manuscript.

3. Manuscripts

3.1. Manuscripts I

Evolution and functional characterization of CAZymes belonging to subfamily 10 of glycoside hydrolase family 5 (GH5_10) in two species of phytophagous beetles

André Busch[1], Grit Kunert[2], David G. Heckel[1] and Yannick Pauchet[1]

[1]Department of Entomology, [2]Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

Manuscript published in

PLoS One

Doi: 10.1371/journal.pone.0184305

PLOS ONE

RESEARCH ARTICLE

Evolution and functional characterization of CAZymes belonging to subfamily 10 of glycoside hydrolase family 5 (GH5_10) in two species of phytophagous beetles

André Busch[1], Grit Kunert[2], David G. Heckel[1], Yannick Pauchet[1]*

1 Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany, **2** Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany

* ypauchet@ice.mpg.de

Citation: 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. https://doi.org/10.1371/journal.pone.0184305

Editor: Jean-Guy Berrin, Institut National de la Recherche Agronomique, FRANCE

Received: June 20, 2017

Accepted: August 21, 2017

Published: August 30, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Max-Planck-Gesellschaft (YP, AB, GK, DGH). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Hemicelluloses, such as xyloglucan, xylan and mannans, consist of a heterogeneous array of plant-derived polysaccharides that form the plant cell wall. These polysaccharides differ from each other in their structure and physiochemical properties, but they share a β-(1,4)-linked sugar backbone. Hemicelluloses can be hydrolyzed by plant-cell-wall-degrading enzymes (PCWDEs), which are widely distributed in phytopathogenic microbes. Recently, it has become apparent that phytophagous beetles also produce their own PCWDEs. Our previous work identified genes encoding putative mannanases belonging to the subfamily 10 of glycoside hydrolase (GH) family 5 (GH5_10) in the genomes of the leaf beetle, *Gastrophysa viridula* (Chrysomelidae, Chrysomelinae; one gene), and of the bean beetle, *Callosobruchus maculatus* (Chrysomelidae, Bruchinae; four genes). In contrast to proteins from other GH5 subfamilies, GH5_10 proteins are patchily distributed within the tree of life and have so far hardly been investigated. We addressed the following questions: Are beetle-derived GH5_10s active PCWDEs? How did they evolve? What is their physiological function? Using heterologous protein expression and enzymatic assays, we show that the *G. viridula* GH5_10 protein is an endo-β-1,4-mannanase. We also demonstrate that only one out of four *C. maculatus* GH5_10 proteins is an endo-β-1,4-mannanase, which has additional activity on carboxymethyl cellulose. Unexpectedly, another *C. maculatus* GH5_10 protein has evolved to use xylan instead of mannans as a substrate. RNAi experiments in *G. viridula* indicate (i) that the sole GH5_10 protein is responsible for breaking down mannans in the gut and (ii) that this breakdown may rather be accessory and may facilitate access to plant cell content, which is rich in nitrogen and simple sugars. Phylogenetic analyses indicate that coleopteran-derived GH5_10 proteins cluster together with Chelicerata-derived ones. Interestingly, other insect-derived GH5_10 proteins cluster elsewhere, suggesting insects have several independent evolutionary origins.

Introduction

The plant's primary cell wall is a complex structure consisting of polysaccharides and proteins that encase and protect growing plant cells. Next to cellulose and pectins, the hemicellulose network—made of polysaccharides such as xyloglucan, xylan and mannans—is one of the main constituents of the plant's primary cell wall [1,2]. The mannan group, which comprises pure mannan, galactomannan and glucomannan, is widely distributed among plants and algae [3] and may be part of the wall of different types of cells and tissues, such as roots, tubers, bulbs and seeds [4]. Mannans may function as seed storage and/or structural components [5,6]. For example, galactomannan, a storage polysaccharide in the endosperm cell wall of legumes, occupies up to 30% of a seed's dry weight [7]. The backbone of the polymer is composed of mannose residues linked together by β-1,4-glycosidic bonds. In galactomannan, this backbone is substituted with α-1,6-linked galactose residues. In contrast to mannan and galactomannan, the backbone of glucomannan is made up of randomly alternating mannose and glucose residues linked together by β-1,4-glycosidic bonds [8].

Endo-β-1,4-mannanase (EC: 3.2.1.78) is a family of so-called plant-cell-wall-degrading enzymes (PCWDEs) that hydrolyze the backbone of mannan polysaccharides into oligosaccharides [9]. These enzymes are widely distributed within the tree of life: they have been found in bacteria, fungi, plants and animals [8]. According to the carbohydrate-active enzymes (CAZy) database (http://www.cazy.org/) [10], endo-β-1,4-mannanases are distributed in several glycoside hydrolase (GH) families, namely, GH5, GH9, GH26, GH44, GH113 and GH134. In metazoans, endo-β-1,4-mannanases have been identified and functionally characterized in bivalves [11,12], gastropods [13,14], Crustacea [15] and a springtail [16]. The common feature of these metazoan mannanases is that they are members of the subfamily 10 of GH5 (GH5_10), according to the current nomenclature of this gene family [17]. This subfamily of GH5 is one of the smallest described to date; as of April 2017, only 28 sequences had been found in the CAZy database. Functionally characterized enzyme members of this subfamily are all endo-β-1,4-mannanases [11,13,14,16]. Genes encoding GH5_10 have also been identified from several bacterial genomes, but to date none has been functionally characterized [18,19]. Neither fungal- nor plant-derived GH5_10 sequences are present in the CAZy database, suggesting that this subfamily of GH5 is absent from these two phyla. During a survey of transcriptomes of several herbivorous beetles member of the Phytophaga clades [20], we identified transcripts encoding GH5_10 putative mannanases in two species of the family Chrysomelidae [21]. This finding indicated that in addition to their ability to break down cellulose and pectins [22–26], some beetles of the Phytophaga clade may also possess the ability to break down mannan polysaccharides. Interestingly, an endo-β-1,4-mannanase has been characterized in a species of Phytophaga beetles—the coffee berry borer, *Hypothenemus hampei*—but it belongs to the subfamily 8 of GH5 (GH5_8) [27,28], suggesting that the ability to break down mannan polysaccharides appeared several times in the evolution of beetles of the Phytophaga clade.

Here we analyze the function and the evolutionary history of GH5_10 putative mannanases encoded by the genome of two chrysomelid beetles with different feeding habits. Larvae and adults of the green dock beetle, *Gastrophysa viridula* (Coleoptera: Chrysomelinae), feed exclusively on the foliage of dock plants (*Rumex* spp.), whereas larvae of the bean beetle, *Callosobruchus maculatus* (Coleoptera: Bruchinae), feed on the galactomannan-rich endosperm of legume seeds. Besides GH5_10 proteins, the genome of *G. viridula*—like other species of the subfamily Chrysomelinae—encodes GH45 and GH48 putative cellulases as well as GH28 pectinases [21,23,29]. In contrast, in *C. maculatus* GH5_10 proteins are only complemented by GH28 pectinases [21,23]. First, we asked whether beetle-derived GH5_10 proteins are active

PCWDEs and how they have evolved. Second, we asked what their physiological function is. We demonstrate that the sole GH5_10 protein of *G. viridula* and one out of four GH5_10 proteins of *C. maculatus* are endo-β-1,4-mannanases. In addition, a second GH5_10 protein of *C. maculatus* has evolved to become an endo-β-1,4-xylanase, which represents the first example of such enzymatic activity in this subfamily of GH5. We show that the genes encoding GH5_10 proteins in these two distantly related chrysomelid beetles share intron positions and phases, and thus have a common origin. Finally, the phylogenetic relationships of these beetle-derived GH5_10 proteins and their counterparts found in other metazoans are quite complex, suggesting that several of the genes in this group of animals may have originated through the acquisition by horizontal gene transfer events from bacterial donors.

Materials and methods

Insect rearing

Gastrophysa viridula adults and larvae were initially collected from broad leaf dock plants (*Rumex obtusifolius*) in the vicinity of Jena, Germany (50˚55'16.4"N 11˚35'14.1"E). No specific permissions were required to collect *G. viridula*. This beetle species is not endangered or protected in any way, and the location where the beetles were collected is a park freely accessible to the public. Collected individuals were brought to the lab and larvae were raised to adulthood. Insects were reared in plastic containers on detached leaves of *R. obtusifolius* grown in a greenhouse. Beetles were allowed to mate and oviposit, and the offspring were used for experiments. Larvae and adults were kept under a light/dark cycle of 16:8 hours at 18˚C and 13˚C, respectively. *Callosobruchus maculatus* originated from a lab culture obtained from Matthew Benton (University of Cologne) and were reared in plastic containers on organic black-eyed peas at room temperature on a lab bench.

Insect cell culture and heterologous expression

Open reading frames (ORFs) were amplified from cDNAs using gene specific primers (S1 Table) designed according to previously described GH5 sequences from *G. viridula* and *C. maculatus* [21]. The forward primer was designed to introduce a Kozak sequence at the beginning of the ORF, and the reverse primer was designed to omit the stop codon. Complementary DNAs (cDNAs) initially generated for RACE-PCR experiments as described by Pauchet and coworkers [21] were used as a template, and PCR reactions were conducted using a high-fidelity Taq polymerase (AccuPrime, Invitrogen). PCR products were cloned into the pIB/V5-His TOPO/TA (Invitrogen), in frame with the coding sequence of a V5-$(His)_6$ epitope. TOP10 competent *E. coli* cells (Invitrogen) were transformed and plated on LB-agar dishes supplemented with 100 μg/ml ampicillin. To select for constructs correctly oriented after ligation into pIB/V5-His TOPO/TA, randomly picked colonies were checked by colony-PCR using the OpIE2 forward primer located on the vector and a gene-specific reverse primer (S1 Table). Positive clones were further cultured in 3 ml DYT-medium containing 100 μg/ml ampicillin. After plasmid isolation using GeneJET Plasmid Miniprep Kit (Thermo Scientific), the ORF of selected clones was fully sequenced in both directions using capillary sequencing to confirm that the ORF had been correctly inserted into the vector and to control that no mutation were introduced during the cloning process. Positive constructs were then transfected in *Sf*9 cells (Invitrogen) using FuGENE HD (Promega) as a transfection reagent. First, successful expression was determined by transiently transfecting three clones per construct in a 24-well plate format. After 72 h, the culture medium was harvested, and successful expression was verified by Western blot using the anti-V5-HRP antibody (Invitrogen). In order to collect enough material for downstream enzymatic activity assays, a single clone per construct was used for

subsequent transfection of insect cells in a 6-well plate format. After 72 h, culture medium was harvested and centrifuged (16,000 x g, 5 min, 4˚C) to remove cell debris; finally the medium was stored at 4˚C until further use. Again, successful expression was verified by Western blot using the anti-V5-HRP antibody.

Agarose diffusion assays

Enzymatic activity of the recombinant proteins was initially assessed using agarose diffusion assays. Agarose (1%) plates were prepared, containing 0.1% substrate (glucomannan, galactomannan and carboxymethyl cellulose) in 40 mM citrate/phosphate buffer pH 5.0. Galactomannan (Megazyme) was derived from Carob pods and had a Galactose to Mannose ratio of 22/78. Glucomannan and carboxymethyl cellulose were both purchased from Sigma Aldrich. Small holes were made in the agarose matrix using cut-off pipette tips, to which 10 μl of the crude culture medium of each produced enzyme was applied. After incubation overnight at 40˚C, activity was revealed by incubating the agarose plate in a 0.1% Congo red solution for 2 h at room temperature followed by a washing step with 1 M NaCl for 30 min at room temperature.

Preparation of primary cell wall from *Rumex obtusifolius* leaves

Plant cell wall was extracted from *R. obtusifolius* leaves according to Feiz *et al.* [30] with slight modifications. Briefly, 32 g of *R. obtusifolius* leaves was blended in 5 mM acetate buffer pH 4.6 and 400 mM sucrose. The plant tissue homogenate was incubated for 30 min at 4˚C while being stirred and then pelleted by centrifugation for 15 min at 1000 x *g* and 4˚C. The pellet was washed twice in 5 mM acetate buffer pH 4.6 containing 0.6 M and 1 M sucrose, respectively. Finally, the pellet was transferred to a 25 μm nylon net (Miracloth) and washed with 6 l of 5 mM acetate buffer pH 4.6. The resulting cell wall was ground in liquid nitrogen and then lyophilized for 48 h. To remove proteins associated with the plant cell wall, 650 mg of lyophilized cell wall material was washed twice in 25 ml of 5 mM acetate buffer pH 4.6 containing 200 mM $CaCl_2$, and was then washed twice in 30 ml 5 mM acetate buffer pH 4.6 containing 1 M NaCl. For each washing step, the cell wall was homogenized by vortexing for 10 min at room temperature and subsequently centrifuged at 4000 x *g* and 4˚C. Subsequently, the protein-free cell wall was washed with 3 l of double distilled water before being lyophilized. Freeze-dried plant cell wall from *R. obtusifolius* was rehydrated in double distilled water, resulting in a 5% stock solution. For thin layer chromatography analyses (see below), 35 μg of the 5% rehydrated PCW was incubated with 30 μl heterologously expressed GH5 from *G. viridula* in a 20 mM citrate/phosphate buffer pH 5.0.

Analysis of hydrolysis reaction products by thin layer chromatography (TLC)

The culture medium of transiently transfected cells was first dialyzed against distilled water at 4˚C for 24 h, using Slide-A-Lyzer Dialysis Cassettes with a 10 kDa cut-off, before being desalted with Zeba Desalt Spin Columns 7 kDa cut-off (both Thermo Scientific), according to the manufacturer's guidelines. Samples were stored at 4˚C until used. Twenty microliter enzyme assays were set up, using 14 μl of dialyzed and desalted crude enzyme extracts mixed with 4 μl of a 1% solution of substrate in a 20 mM citrate/phosphate buffer pH 5.0. The following substrates were tested: carboxymethyl cellulose, beechwood xylan (both Sigma Aldrich), glucomannan, galactomannan and xyloglucan (all Megazyme). Additionally, the manno-oligomers, D-(+) tetraose to D-(+) hexaose (Megazyme) were tested at a final concentration of 250 ng/μl. Samples were then incubated overnight at 40˚C. Finally, 15 μl of the reaction was

applied to TLC plates (silica gel 60, Merck) and enzymatic breakdown products were separated using the following mobile phase: butanol/glacial acetic acid/water (2:1:1). Breakdown products were revealed by spraying the TLC plates with 0.2% (w/v) orcinol in methanol/sulfuric acid (9:1) followed by heating until reaction products appeared. The reference standard contained 2 μg each of mannose, mannobiose, mannotriose, mannotetraose and mannopentaose (all Megazyme) or 2 μg each of glucose, cellobiose, cellotriose, cellotetraose and cellopentaose (all from Sigma-Aldrich) or 2 μg each of xylose (Sigma-Aldrich), xylobiose and xylotriose (both Megazyme), according to the substrate tested.

Temperature optimum and pH optimum

To test the temperature optima, dialyzed and desalted crude enzyme extracts were incubated with 0.5% (w/v) galactomannan (GVI1), or galactomannan and carboxymethyl cellulose in parallel (CMA3), or beechwood xylan (CMA2) in 20 mM citrate phosphate buffer (pH 5.0) at different temperatures ranging from 20˚C to 80˚C in steps of 10˚C. In detail, each enzyme assay was performed with 24 μl crude enzyme extract, 30 μl of 1% (w/v) substrate solution and 6 μl of 20 mM citrate phosphate buffer pH 5.0. Negative controls were carried out with 24 μl of distilled water instead of enzyme. The enzymatic activity was assayed at 40˚C for 5 min (GVI1), 2.5 h (CMA3 against GalM), 16 h (CMA3 against CMC) and 16 h (CMA2). The amount of reducing sugars produced in these reactions was measured using the dinitrosalicylic acid (DNS) method according to Kirsch and co-workers [23]. To test for pH optima, dialyzed and desalted crude enzyme extracts were incubated with their respective substrate as described above and assayed in 20 mM citrate phosphate buffers ranging from pH 2.0 to 9.0 as well as in 20 mM sodium carbonate buffer pH 10.0. The amount of reducing sugars produced in these reactions was measured using the dinitrosalicylic acid (DNS) method as described above. Each reaction was carried out in triplicate.

Preparation of double-stranded RNA and off-target prediction

Primers for RNA interference (RNAi) experiments were designed for *G. viridula* GH5 (GVI1) and GFP used as controls, yielding a 300 bp fragment and a 379 bp fragment, respectively (S1 Table). To predict potential off-target effects, the sense and anti-sense RNA strands were diced *in silico* into all possible 21 bp fragments using an in-house algorithm. The resulting siRNAs were searched against our *G. viridula* larval gut transcriptome [21], using previously described parameters [31]. A siRNA was considered off-target if the resulting hit was equal to or higher than 21 bp by allowing one mismatch. Gene fragments were amplified from sequenced recombinant plasmids containing GVI1 or GFP. The amplicons were gel-purified using Zymoclean Gel DNA recovery Kit (Zymo Research). To obtain double-stranded RNA (dsRNA), the purified PCR product was used as a template for *in vitro* transcription using the MEGAscript RNAi kit (Ambion), following the manufacturer's instructions. To remove residual DNA contamination, the resulting dsRNA was nuclease-digested using TURBO™ DNase (Thermo Scientific), and then purified and recovered in 150 μl injection buffer (3.5 mM Tris–HCl, 1 mM NaCl, 50 nM Na2HPO4, 20 nM KH2PO4, 3 mM KCl, 0.3 mM EDTA, pH 7.0). The quantity of dsRNA was estimated using a spectrophotometer (NanoDrop ND-1000, Peqlab Biotechnology), and its quality was assessed by gel electrophoresis.

Injection of dsRNA and assessment of RNAi efficiency

Early second-instar *G. viridula* larvae were injected dorsally with 50 nl (150 ng) of target dsRNA into the metathorax, using a Nanoliter 2010 Injector (World Precision Instruments) attached to a three-dimensional micromanipulator, and were then put onto fresh *R.*

obtusifolius leaves. To record weight gain and mortality, five animals per replicate were injected with a total of six replicates for each target gene. To analyze gene expression and enzymatic activity, three animals per replicate were injected with a total of six replicates for each target gene. In addition to larvae injected with dsRNA targeting GFP, a non-injected control was also included. For quantitative PCR and enzymatic activity analyses, larvae were collected at days 1, 4 and 8 post injection. Whole larvae were crushed in liquid nitrogen and separated in half. One aliquot was used for total RNA preparation, the other for protein extraction.

Total RNA was isolated using innuPREP RNA Mini Kit (Analytik Jena), following the manufacturer's protocol. The resulting RNA was then subjected to DNase digestion (Ambion), and its quality was subsequently checked using the RNA 6000 Nano LabChip kit on a 2100 Bioanalyser (both Agilent Technologies). Total RNA was used as a template to synthesize cDNAs using the Verso cDNA synthesis kit (Thermo Scientific). The resulting cDNA samples were then used for real-time qPCR experiments, which were performed in 96-well hard-shell PCR plates on the CFX Connect Real-Time System (both Biorad). All reactions were carried out using the 2-Step QPCR SYBR Kit (Thermo Scientific), following the manufacturer's instructions. Primers were designed using Primer3 (version 0.4.0) (S1 Table). The specific amplification of each transcript was verified by dissociation curve analysis. A standard curve for each primer pair was determined in the CFX Manager (version 3.1) based on Cq-values (quantitation cycle) of qPCRs run with a dilution series of cDNA pools. The efficiency and amplification factors of each qPCR, based on the slope of the standard curve, were calculated using an integrated efficiency calculator of the CFX manager software (version 3.1). The sequence of the transcript encoding ribosomal protein S3 (*RPS3*), extracted from our *G. viridula* larval gut transcriptome [21], was used as a reference for all qPCR experiments, and the abundance of GVI1 transcripts was expressed as RNA molecules per 1000 RNA molecules of *RPS3*.

To directly compare GH5 transcript abundance to GH5 enzymatic activity in RNAi-treated *G. viridula*, crushed and frozen material was suspended in 40 mM citrate/phosphate buffer at pH 5.0 containing a protease inhibitor cocktail (Complete EDTA-free, Roche). Then, the samples were centrifuged (10 min, 16000xg, 4˚C), and the supernatant was collected and stored at 4˚C until further use. Protein concentration was estimated by Bradford Protein Assay (Bio-Rad). Enzymatic activity assays were carried out using the DNS method as described above, using 2 μg of extracted proteins in the reaction. Alternatively, 0.5 μg total extracted proteins were prepared for zymogram analysis by diluting the sample in Laemmli buffer without any reducing agent. Samples were run on a 12.5% SDS-PAGE gel containing 0.1% (w/v) galactomannan. Electrophoresis was carried out at 4˚C using pre-chilled running buffer. Gels were then washed three times in a 2.5% Triton X-100 solution for 15 min, each at 4˚C, before being equilibrated in the reaction buffer (50 mM citrate/phosphate buffer pH 5.0) for 16 h at 4˚C, followed by a 1 h incubation at 40˚C. The gels were then incubated in a 0.1% (w/v) Congo red solution before being destained in 1 M NaCl until pale activity zones appeared against a dark red background.

Two life history traits were recorded after larvae were injected with dsRNA. First, larvae (we used groups of five insects per replicate, six replicates in total) were weighed on day 1and day 8 post injection. Then, growth rate was calculated using the formula "*Growth rate* = Log_{10}*(Final weight)*-Log_{10}*(Initial weight)/ Time (days)*". Finally, mortality was recorded at the end of the experiment.

Tissue-specific gene expression

Late-instar *G. viridula* larvae, actively feeding on leaves of *R. obtusifolius*, as well as late-instar *C. maculatus* larvae, actively feeding inside black-eyed peas, were used for RNA extraction.

Larvae were cut open from abdomen to head, and the complete gut was removed and stored separately from the rest of the body. Dissection and storage were carried out in RL solution (Analytik Jena). Three biological replicates were sampled, each containing three larvae. RNA extraction, generation of cDNAs and subsequent real-time qPCR experiments were performed as described above. Primers used for these experiments are listed in S1 Table.

Statistical analyses

If not otherwise stated data were analyzed in R version 3.2.0 [32]. Statistical analyses of gene expression over time were performed as follows: The influence of GVI1 RNAi treatment (iGH5) over time RNAi treatment and time used as categorical explanatory variables on GVI1 transcript abundance was investigated using the generalized least squares method (gls from the nlme library [33]) to account for the variance heterogeneity among the residuals. The varIdent variance structure was used, with a different variance for the combination of treatment and time (varIdent (form = ~1|combination of [treatment and time])). The influence of the explanatory variables was determined by sequentially removing explanatory variables starting with the full model and comparing the simpler model to the more complex one, using a likelihood ratio test [34]. Differences between factor levels were determined by factor level reduction [35]. The influence of RNAi treatment on the enzyme activity over time was analysed with a two-way ANOVA. The Tukey HSD test was performed in order to find differences between the groups. To compare weight gain over time in RNAi-treated larvae, we calculated the relative growth rate for the period of 8 days and analyzed the data in SigmaPlot version 11.0 (Systat Software) using a one-way ANOVA. Differences in mortality were analyzed using the equality of proportions–test. Differences in tissue-specific gene expression were analyzed with paired t-tests again in SigmaPlot Version 11.0.

Gene structure determination

The GVI1 ORF was amplified by PCR from genomic DNA using gene-specific primers (S1 Table). Genomic DNA was prepared from a single *G. viridula* male beetle using the QIAamp DNA micro kit (Quiagen), following the manufacturer's instructions. The PCR product was then cloned into the pCR4 TOPO TA vector (Invitrogen), followed by the transformation of TOP10 competent *E. coli* cells (Invitrogen). Cells were then plated on LB-agar dishes supplemented with 100 μg/ml ampicillin. To select for constructs harboring the sequence of interest, randomly picked colonies were checked by PCR using M13 forward and reverse primers. Plasmid DNA was prepared from positive clones using the GeneJET Plasmid Miniprep Kit (Thermo Scientific). The sequence of the GVI1 ORF was deduced from three independent clones after capillary sequencing. Sequences corresponding to the genes encoding CMA1, CMA2, CMA3 and CMA4 were retrieved from a draft genome assembly of *C. maculatus* made publicly available (http://www.beanbeetles.org/genome/). The intron/exon structure was determined for each gene using Splign [36].

Amino acid alignment and phylogenetic analyses

All sequences corresponding to GH5_10 proteins present in the carbohydrate-active enzymes (CAZy) database [37] as of February 1, 2017, were retrieved. Due to the paucity of GH5_10 sequences available, extra searches were conducted in bacterial and fungal genomes available through the genome portal of the Joint Genome Institute (http://jgi.doe.gov/) as well as in transcriptome shotgun assemblies available at Genbank (https://www.ncbi.nlm.nih.gov/genbank/tsa/). A description of the sequences can be found in S2 Table. All obtained sequences were analyzed for the presence of a signal peptide and extra protein domains other than the

GH5 domain using InterProScan version 61.0. Once such features were identified, they were removed from the collected protein sequences, and only the GH5 domain was conserved for amino acid alignment. Amino acid alignments were carried out using MUSCLE version 3.7 on the Phylogeny.fr web platform (http://www.phylogeny.fr) [38], and were inspected and corrected manually when needed. Maximum-likelihood-inferred phylogenetic analyses were conducted in MEGA6 [39]. The best model of protein evolution was determined in MEGA6 using the 'find best DNA/protein models' tool. The best model was the 'Le and Gascuel' (LG) model, incorporating a discrete gamma distribution (shape parameter = 5) to model differences in evolutionary rates among sites (+G) and a proportion of invariable sites (+I). The robustness of each analysis was tested using 1,000 bootstrap replicates.

Results

Characterization of *G. viridula* and *C. maculatus* GH5_10 proteins reveals distinct enzymatic activities

Full-length amplicons of target GH5_10 transcripts were cloned into a pIB-V5/His TOPO vector and transiently expressed in insect *Sf*9 cells. Validation that GH5_10 proteins were successfully expressed and secreted into the culture medium was made by Western blot; heterologous proteins had an apparent molecular weight of 45 to 55 kDa close to their theoretical expected size ranging from 41.7 to 43.2 kDa (Fig 1A). To test whether these GH5_10 proteins were enzymatically active, we initially analyzed crude enzyme extracts on agarose diffusion plates containing various plant cell wall polysaccharides as substrates (Fig 1B). *Gastrophysa viridula* GH5-1 (GVI1) and *C. maculatus* GH5-3 (CMA3) exhibited activity halos on plates containing galactomannan (GalM) and glucomannan (GluM). In addition, CMA3 showed activity halos on plates containing carboxymethyl cellulose (CMC).

To further analyze the enzymatic properties of beetle-derived GH5_10 proteins, we used in-tube assays with an array of cellulosic and hemicellulosic poly- and oligosaccharides, and analyzed them by TLC (Fig 2). GVI1 exhibited activity against GalM and GluM (Fig 2A and 2B). GalM breakdown products consisted mainly of trimers and larger oligomers and, to a lesser extent, monomers and dimers of mannose. GluM breakdown products seemed to be trimers and tetramers and, to a lesser extent, dimers and monomers (Fig 2A and 2B). Compared to GalM breakdown products, however, these oligomers appear to be far less resolved on TLC. Most likely, this discrepancy in the resolution of breakdown products lies in the chemical nature of both substrates. GalM is a polysaccharide consisting of a pure mannose backbone decorated with evenly distributed galactose moieties, whereas GluM is a straight-chain polysaccharide consisting of a backbone that alternates unevenly between mannose and glucose moieties with occasional branching. Thus, we believe that the heterogeneous structure of GluM leads to inconsistently sized breakdown products which appear as smears on TLC. We then tested the ability of GVI1 to cleave several mannan oligomers (Fig 2). GVI1 cleaved mannohexaose into mannotriose (Fig 2F), and mannopentaose into mannotriose and mannobiose (Fig 2G). The smallest mannan oligomer that GVI1 could cleave was mannotetraose, resulting in the breakdown products mannotriose, mannobiose and mannose (Fig 2H).

Like GVI1, CMA3 exhibited activity against galactomannan and glucomannan (Fig 2A and 2B), but CMA3 was also able to break down CMC (Fig 2C). The main breakdown products accumulating after GalM and CMC degradation are the corresponding trioses and the larger oligomers. As we observed for GVI1, GluM breakdown products did not resolve very well on TLC. The smallest oligomer that CMA3 was able to cleave was mannohexaose, releasing mannotriose (Fig 2F). The results obtained for GVI1 and CMA3 on TLC confirmed the activity

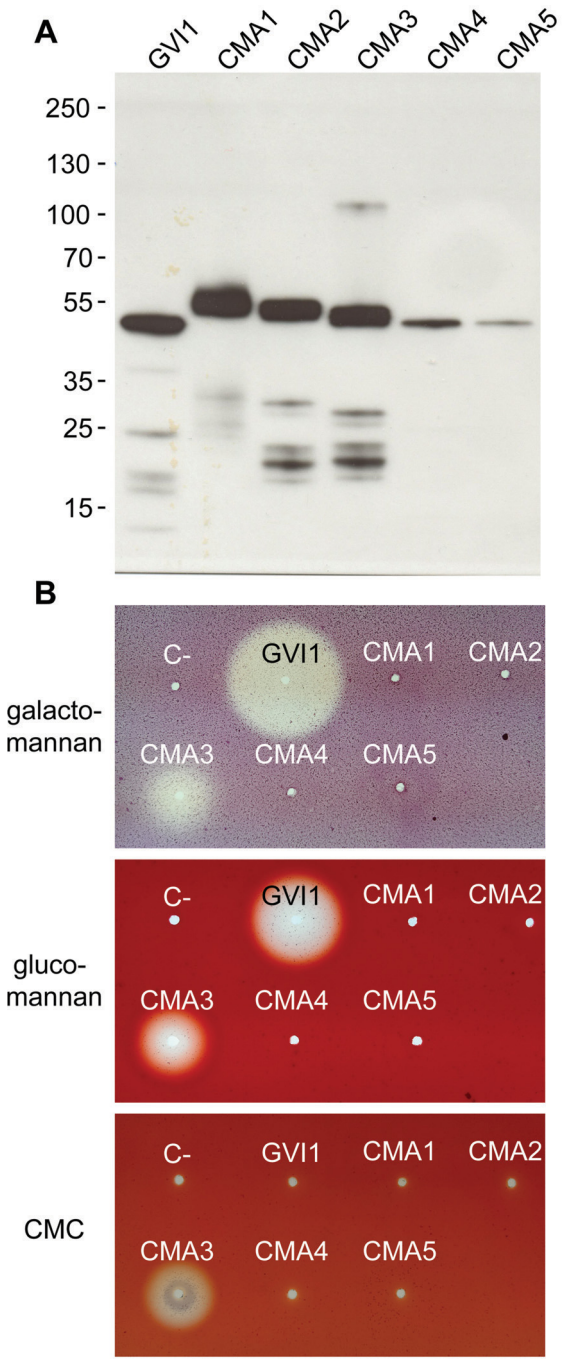


Fig 1. Heterologous expression of GH5_10 proteins from *G. viridula* and *C. maculatus* in *Sf*9 insect cells. (A) GH5_10 cDNAs cloned into an expression vector in frame with a V5/$(His)_6$ epitope were transfected into *Sf*9 cells. The culture medium of transfected cells was collected 72 hours post transfection and samples were subjected to Western blot. An anti-V5-HRP antibody was used for detection and the blot was revealed using chemiluminescence. Molecular weight markers are indicated next to the Western blot. (B) The culture medium of transfected cells was applied to agarose plates containing 0.1% substrate in McIlvain buffer pH

5.0, and plates were incubated for 16 hours at 40˚C. Activity halos were revealed after staining with Congo red.

https://doi.org/10.1371/journal.pone.0184305.g001

observed in agarose plate assays. Our data strongly indicate that GVI1 and CMA3 are endo-β-1,4-mannanases, with CMA3 also acting as endo-β-1,4-glucanase.

Unexpectedly, TLC experiments allowed us to detect a second enzymatically active GH5_10 protein in *C. maculatus* (CMA2), which was able to degrade beechwood xylan (Fig 2D). The breakdown products generated by CMA2 were xylotriose and larger oligomers of xylan, which suggests that this enzyme acted as an endo-β-1,4-xylanase. GH5 family members harboring xylanase activity have been described in bacteria [40,41] and, recently, in cerambycid beetles [25,26]. However, these xylan-degrading GH5 enzymes were encoded by distinct GH5 subfamilies (GH5_2, _4 and _21) [17]. CMA2 represents the first example of a GH5_10 protein harboring endo-β-1,4-xylanase activity.

None of the tested heterologously expressed proteins showed activity against xyloglucan (Fig 2E). No activity either on plates or on TLC was observed for *C. maculatus* GH5-1, -4, -5 (CMA1, CMA4, CMA5) on any of the substrates tested. An alignment of the amino acid sequences including all coleopteran-derived GH5_10 proteins together with two additional sequences of proteins with known crystal structures [11,42] showed that the two catalytic glutamate residues are conserved, and confirmed that no dramatic substitutions of active site residues occurred between the two reference sequences and the sequences derived from beetles (S1 Fig). These patterns may indicate that CMA1, CMA4 and CMA5 are active enzymes but that their substrate has not yet been discovered or, alternatively, that they lost their activity due to mutations in other functionally important sites.

Investigation of optimal pH values and temperatures for coleopteran GH5_10 enzymes

The enzymatic performance of GVI1 was monitored using galactomannan as a substrate. GVI1 performed best at acidic pH values with an optimum around pH 5.0 (Fig 3A) and a temperature optimum close to 50˚C (Fig 3B). As CMA3 was the only enzyme we discovered exhibiting activity against three substrates—namely, GalM, GluM and CMC—we chose to test CMA3 against GalM and CMC in parallel. The optimal pH value for CMA3 using GalM as a substrate was 5.0, and the corresponding optimal temperature was around 40˚C (Fig 3A and 3B). The optimal pH value for CMA3 tested against CMC was close to 4.0 and the corresponding optimal temperature, around 40˚C (Fig 3A and 3B). The enzymatic performance of CMA2 was monitored using beechwood xylan as a substrate. The pH optimum for CMA2 was determined to be around 6.0 and the optimal temperature was approximately 50˚C. In summary, each enzyme analyzed performed best under acidic conditions, which correlates well to the pH conditions of the gut lumen in related beetle species [43].

Tissue-specific gene expression and gene silencing of the *G. viridula* GH5_10 gene

To learn where the genes encoding GH5_10 family members are expressed in *G. viridula* and *C. maculatus*, we performed quantitative RT-PCR on midgut tissue and on the rest of the body. Transcripts encoding all GH5_10 proteins in both species are significantly more abundant (statistical values see S3 Table) in the midgut tissue, whereas almost no transcripts were detected in the rest of the body, reinforcing the fact that these GH5_10 proteins have a digestive function (S2 Fig).

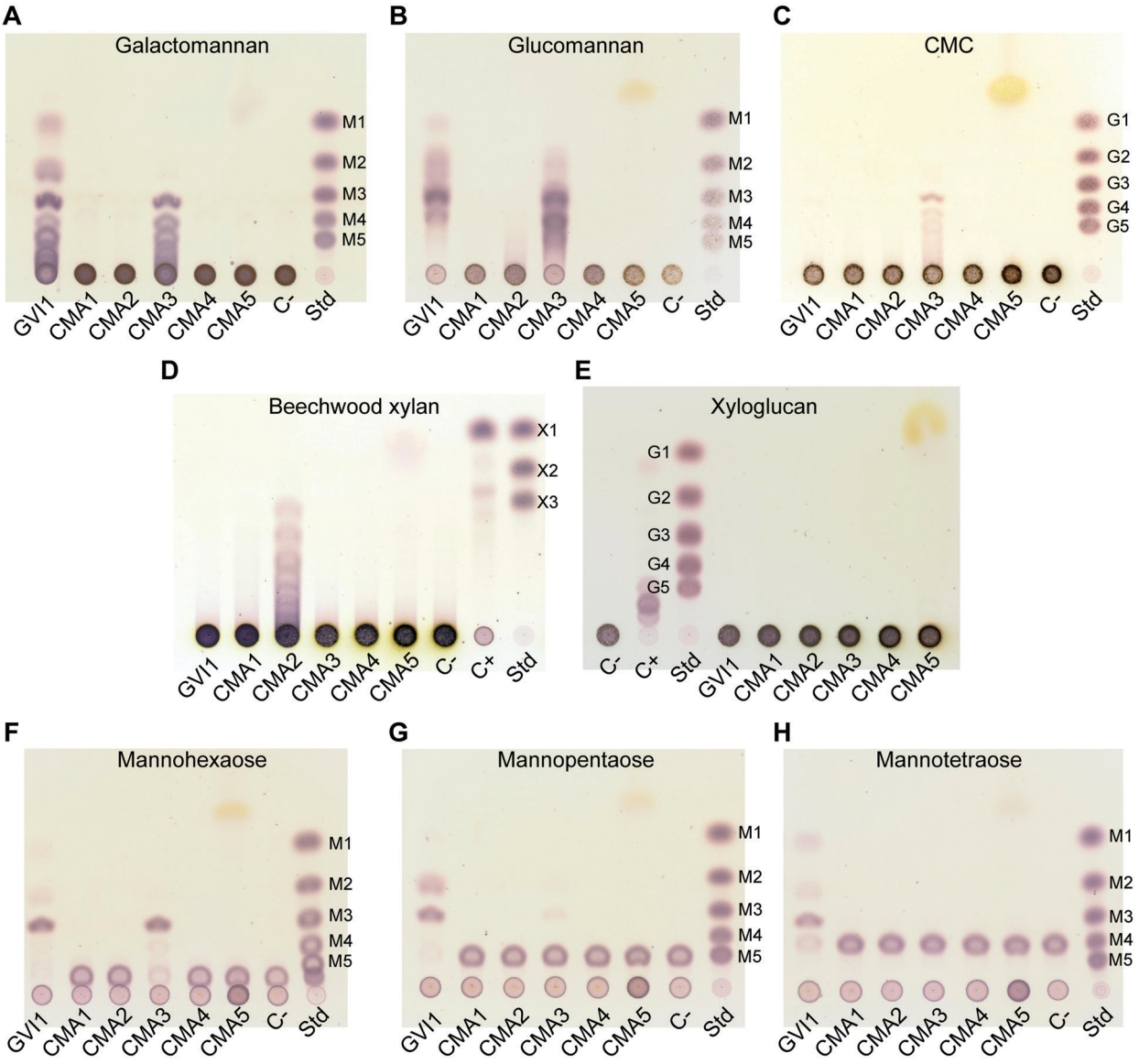


Fig 2. Thin-layer chromatograms of beetle GH5_10 assays against a range of plant cell wall polysaccharides and mannan oligomers. (A) Heterologously expressed *G. viridula* and *C. maculatus* GH5_10 proteins were incubated with galactomannan. GVI1 releases mannose, mannobiose, mannotriose and larger oligomers. CMA3 releases mannotriose and larger oligomers. (B) The same proteins incubated with glucomannan. GVI1 and CMA3 release a range of oligomers, which proved difficult to resolve on TLC. (C) The same proteins incubated with carboxymethylcellulose (CMC). CMA3 releases cellotriose and larger oligomers. (D) The same proteins incubated with beechwood xylan. CMA2 releases xylotriose and larger oligomers. (E) The same proteins incubated with xyloglucan. None of the proteins showed activity against this substrate. (F) The same proteins incubated with mannohexaose. Both GVI1 and CMA3 release mannotriose. (G) The same proteins incubated with mannopentaose. GVI1 releases mannotriose and mannobiose. (H) The same proteins incubated with mannotetraose. GVI1 releases mannose, mannobiose and mannotriose. Standards (Std) used are mannose to mannopentaose (M1 to M5), glucose to cellopentaose (G1 to G5), xylose to xylotriose (X1 to X3). A negative control was introduced (C-) to which no enzyme was added. A positive control (C+), which is composed of a commercial cellulase preparation from *Trichoderma reesei* incubated with the corresponding substrates, was included in the TLCs of xylan and xyloglucan activity assays.

https://doi.org/10.1371/journal.pone.0184305.g002

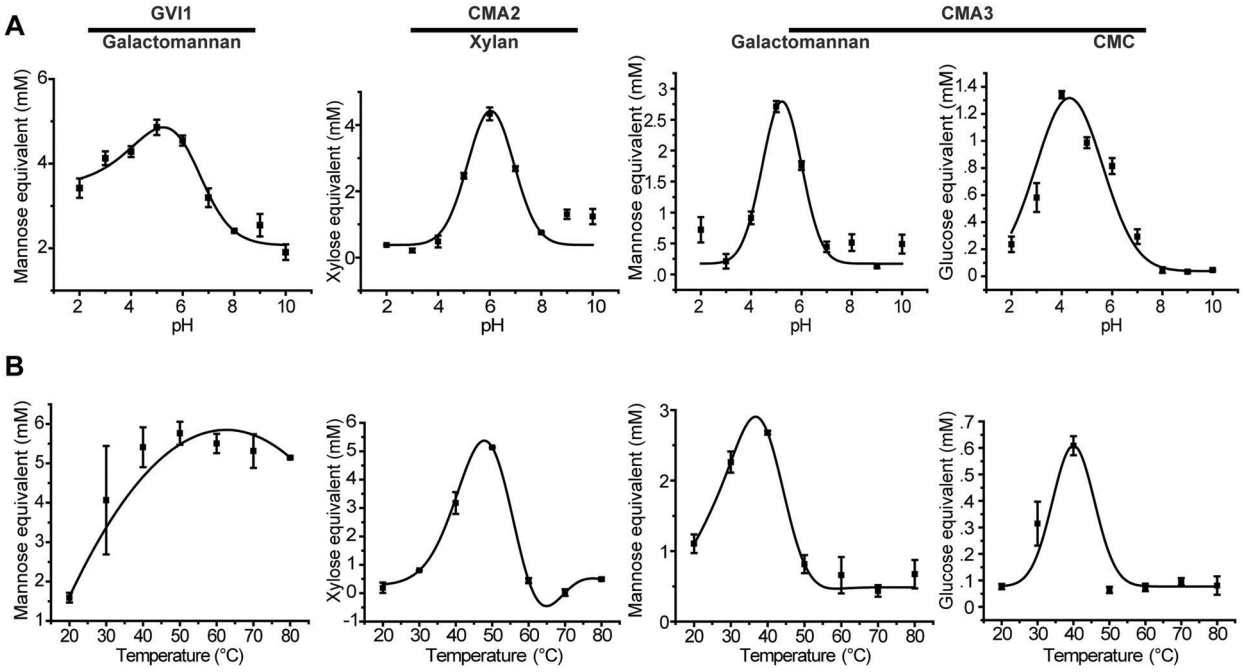


Fig 3. Determining the optimal pH values and temperatures of the enzymatically active GH5_10 proteins. (A) GVI1, CMA2 and CMA3 were incubated with their respective substrates at various pH values, ranging from 2.0 to 10.0. (B) The same proteins were incubated with their respective substrates at various temperatures, ranging from 20 to 80˚C. The amount of reducing sugars released was determined by DNS assay and converted into millimolar (mM) of sugar monomer equivalent. The results are the means of three independent replicates ±SEM. The substrates used were galactomannan for GVI1 and CMA3, beechwood xylan for CMA2 and carboxymethylcellulose (CMC) for CMA3.

https://doi.org/10.1371/journal.pone.0184305.g003

To investigate how important GH5_10 are for the biology of these beetles, we used dsRNA-mediated silencing of the expression of the gene encoding GVI1 in *G. viridula* and analyzed the subsequent genotype and phenotype. We decided to perform RNAi experiments only with G. viridula. In fact RNAi is not really possible to perform with C. maculatus mainly due to the biology of this beetle species. As mentioned earlier, C. maculatus larvae develop inside legume seeds. One would have to remove the larvae out of the seeds to inject the dsRNA and then put them back into a legume seed which is practically not possible.

First, we examined gene expression levels of GVI1 at several time points after the injection of dsRNA (Fig 4A). We managed to significantly knock down the expression of the GVI1 gene (iGH5) compared to insects injected with dsRNA targeting GFP as control (iGFP) (likelihood ratio = 59.634, $p < 0.001$). The changes of the expression levels over time differed between treatments (likelihood ratio = 21.114, $p<0.001$). More precisely, we were able to reduce the expression of the GVI1gene up to 96.6% at day 4 post injection, 70.8% at day 8 post injection and 64.7% in adults compared to iGFP control animals. In general, our knockdown of the expression of the GVI1 gene using dsRNA proved to be stable and lasted through the adult stage (Fig 4A). Second, we measured the mannanase activity of guts dissected from iGH5 larvae at several time points. We found that levels of mannanase activity in iGH5 insects were significantly reduced compared to iGFP and NIC control insects (Fig 4B) ($F = 71.361$, $p< 0.001$). The changes of enzymatic activity over time differed between treatments ($F = 3.499$, $p = 0.014$) and was reduced to 73.96% at day 4 post injection, 70.32% at day 8 post injection and 57.93%

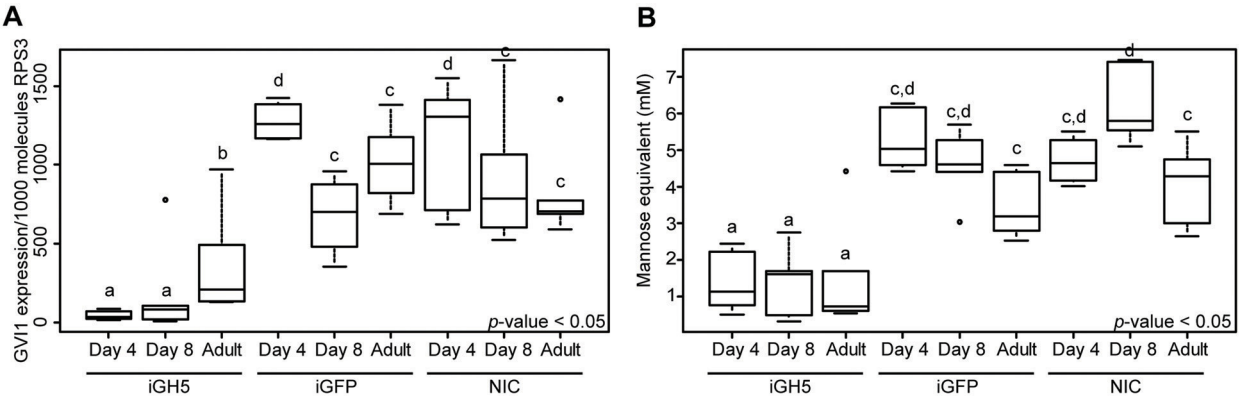


Fig 4. Knockdown of the expression of the gene encoding GVI1 by RNA interference. Early second-instar larvae of *G. viridula* were injected with double-stranded RNA (dsRNA), targeting GVI1 (iGH5) or targeting GFP (iGFP) as a control. A non-injected control (NIC) was also included. Larvae were collected on days 4 and 8 post injection. Newly emerged adults were also collected. Groups of three insects (six replicates per treatment) were snap-frozen in liquid nitrogen before being ground into a fine powder. Half of the powder was used for total RNA preparation and subsequent quantitative RT-PCRs, and the other half was used in enzyme assays. (A) The expression of the gene encoding GVI1 was assessed in the different treatments by quantitative RT-PCR. The gene expression is given as copy number per 1000 RNA molecules of RPS3. (B) Quantification of the mannanase activity in the same treatments. The amount of reducing sugars released was determined by DNS assay and converted into millimolar (mM) of mannose. The different letters on top of each box plot indicate significant differences (P < 0.05). For details on the statistics, please refer to the Materials and Methods section.

https://doi.org/10.1371/journal.pone.0184305.g004

in adults compared to iGFP control insects. Again, the reduction of mannanase activity remained stable in the gut of iGH5 insects from the larval to the adult stage (Fig 4B).

Although the reduction of mannanase activity in iGH5 insects was substantial and correlated with the knockdown of the expression of the gene encoding GVI1, it did not reach the same level (that is, ca. 95% reduction). Therefore we asked whether GVI1 is the only mannanase present in the gut of *G. viridula*. To answer this question, we performed zymogram analyses using galactomannan as a substrate, and compared iGH5 to iGFP and non-injected control insects (S3 Fig). A single band harboring mannanase activity could be seen in all samples. The intensity of the band in iGH5 protein samples was strongly reduced compared to the intensity of the band from the control insects, indicating (i) that GVI1 is the sole endo-β-1,4-mannanase present in the gut fluid of *G. viridula*; and (ii) that there is still a non-negligible amount of this enzyme even after RNAi.

We then monitored two life history traits (growth rate and mortality) and compared them between iGH5 insects and both iGFP and non-injected control insects over several days post injection (Fig 5). We chose day 8 post injection for our analysis as this was the last day larvae actively fed. Although our data suggested iGH5 insects tend to grow more slowly compared to control insects, no statistically significant differences could be documented (Fig 5A) ($F = 1.875$, $p = 0.188$). Additionally, we saw no significant differences in the mortality of larvae injected with GH5 dsRNA compared to control larvae (Fig 5B, $\chi^2 = 0.423$, $p = 0.809$). In summary, although we managed to knock down the expression of the gene encoding GVI1 and reduce levels of mannanase activity, we observed no differences in growth and mortality. As a result, we believe that GVI1 may not have a primary digestive function but is more likely an accessory enzyme.

Significance for *G. viridula* to express a mannanase

Taking into account that the amount of mannan polysaccharides in the primary cell wall of plant leaves is usually comprised between 2 and 5% [44], we wondered whether, the action of

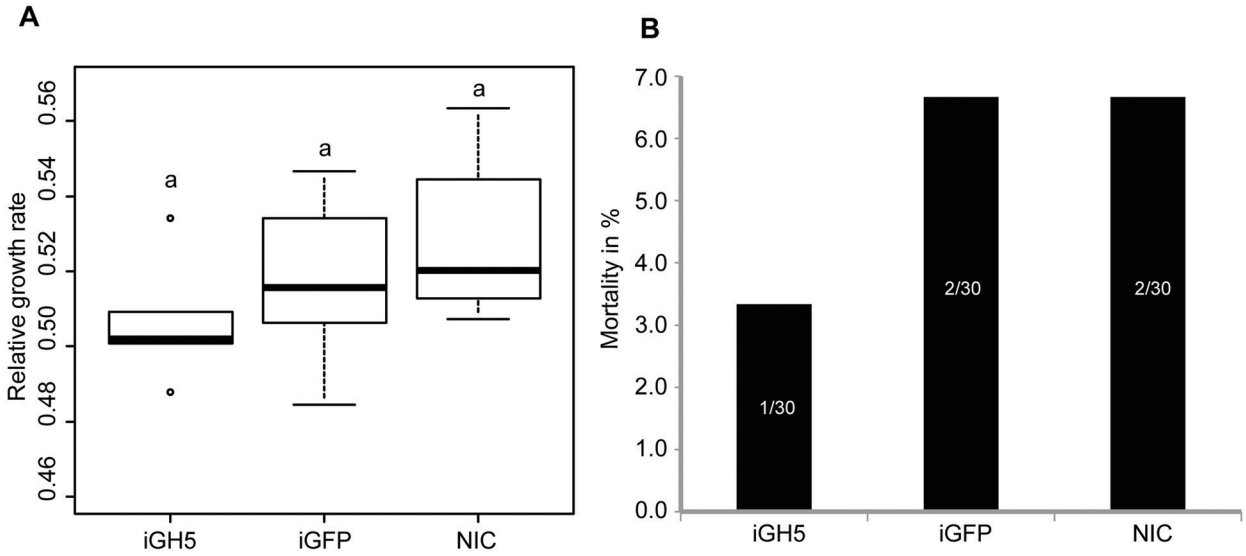


Fig 5. Effects of knocking down the expression of the gene encoding GVI1 on growth rate and mortality. Early second-instar larvae of *G. viridula* were injected with double-stranded RNA (dsRNA) targeting GVI1 (iGH5) or targeting GFP (iGFP) as a control. A non-injected control (NIC) was also included. Larvae were collected at days 4 and 8 post injection. Newly emerged adults were also collected. (A) Groups of five insects (six replicates per treatment) were weighed on day 1 and day 8 after they were injected with dsRNA, and growth rates were calculated. A one-way ANOVA statistical test was applied to the data. (B) The number of dead larvae per treatment was recorded during the eight days of the experiment. Mortality data were analyzed using the equality of proportions–test.

https://doi.org/10.1371/journal.pone.0184305.g005

GVI1 could release mannan oligomers from the cell wall of *Rumex* leaves. To test this hypothesis, we isolated primary cell wall from leaves of *R. obtusifolius* free of proteins and incubated it with GVI1 heterologously expressed in *Sf*9 cells; then we visualized the results of this experiment on TLC (S4 Fig). We observed the presence breakdown products in cell wall samples treated with GVI1, whereas no oligomers could be detected in cell wall samples incubated alone, indicating that the primary cell wall of *R. obtusifolius* leaves possesses polysaccharides that can be broken down by GVI1 in the gut of *G. viridula* beetles.

Phylogenetic analysis and evolution of GH5_10 proteins

To clarify the evolutionary history of GVI1 and CMA1–4 (CMA5 was excluded from the analysis as we believe it is an allele of CMA4), we reconstructed their molecular evolution in a phylogenetic analysis. We performed extensive searches of GH5_10 genes in various databases, including the CAZy database (http://www.cazy.org; [37]), as well as in publicly available genome and transcriptome assemblies. We identified GH5_10 genes in several other insects (Zygentoma, Archaeognatha, Ephemeroptera) and also in Collembola, Crustacea, Chelicerata of the family Oribatidae, mollusks and bacteria (See S2 Table). According to current data, GH5_10 genes seem to be absent in plants and fungi. Amino acid sequences collected from our search were first aligned with each other and a maximum-likelihood-inferred phylogenetic analysis was performed, showing that the coleopteran-derived GH5_10 proteins clustered in a highly supported clade together with Oribatidae (Chelicerate)-derived proteins (Fig 6). Interestingly, other insect-derived GH5_10 proteins do not cluster together with the coleopteran proteins but form a well-supported clade together with crustacean- and collembolan-derived GH5_10 proteins. This heterogeneous distribution of insect GH5_10 proteins may hint at

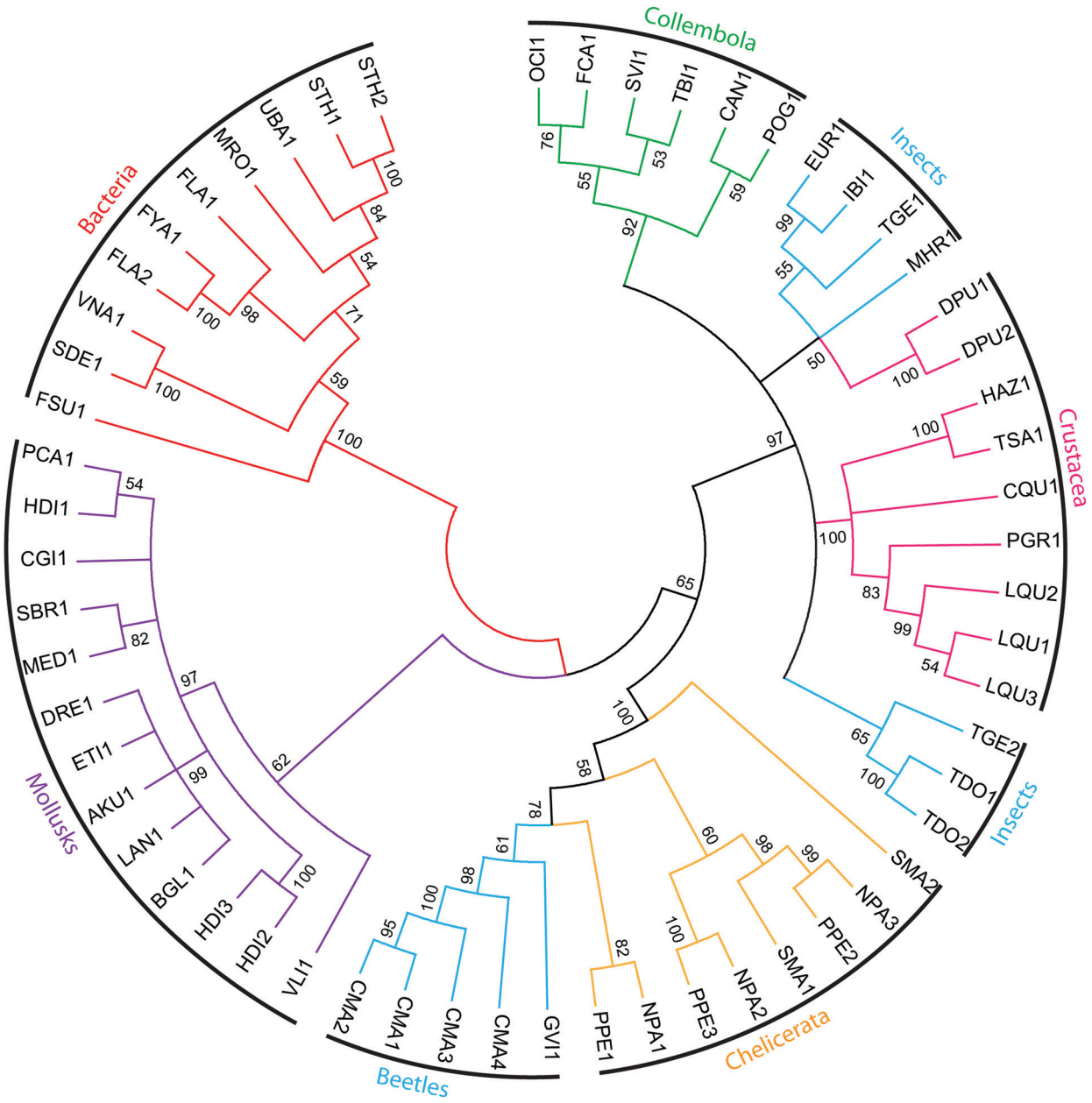


Fig 6. Phylogenetic relationships among beetle GH5_10 proteins and other animals and bacteria. A maximum-likelihood-inferred phylogeny is shown which compares the predicted amino acid sequences of the GH5_10 proteins from *G. viridula* and *C. maculatus* described here with their other animal and bacterial counterparts. Bootstrap support values (1000 replicates) are indicated at corresponding nodes. When the bootstrap support value of a given node was below 50, the corresponding node was condensed. Details of the sequences used for the analyses as well as accession numbers are provided in the electronic supplementary material, S2 Table. Branches in blue correspond to insect proteins; branches in red to bacterial proteins; branches in purple to mollusk proteins; branches in orange to chelicerate proteins; branches in green to collembolan proteins; and branches in pink to crustacean proteins.

https://doi.org/10.1371/journal.pone.0184305.g006

several independent evolutionary origins for this gene family in insects, indicating the likelihood of several potential horizontal gene transfer events from bacterial donors.

We investigated beetle-derived GH5_10 encoding genes in more detail by examining intron-exon structures (S5 Fig). We investigated three full-length (GVI1, CMA3, CMA4) and two partial genomic sequences (CMA1, CMA2). For that, we PCR-amplified the gene encoding GVI1 from genomic DNA, and we acquired the structure of the CMA GH5_10 genes from a draft genome sequence. Two introns were identified in the gene encoding GVI1. The position of these two introns is conserved in the genes encoding GVI1 and the CMA GH5_10 genes. We also found extra introns in the CMA sequences. Not only do the gene encoding GVI1 and the CMAs GH5_10 genes share two intron positions, but the phase of these introns is also conserved, suggesting that the most recent common ancestor of *G. viridula* and *C. maculatus* may have possessed one or several GH5_10 genes in its genome. Also, the presence of introns in all target genes confirmed that these genes are endogenous to the beetles' genomes.

Discussion

We previously reported that the transcriptomes of two phytophagous beetles, *G. viridula* and *C. maculatus*, harbor transcripts encoding GH5_10 putative mannanases [21]. Here, we demonstrated that at least some of these GH5_10 proteins are indeed enzymatically active on plant cell wall polysaccharides, implying that they have a digestive function. According to our data, the gene encoding GVI1 is expressed in the gut and the corresponding protein is secreted into the gut lumen. The same is also true for the GH5_10 genes and corresponding proteins found in *C. maculatus*. A transcriptome analysis of *C. maculatus* indicated that the transcripts corresponding to CMA3 and CMA4 were expressed specifically in larval gut tissue [45]. In addition, the corresponding proteins, as well as CMA2, were identified in a proteome analysis of *C. maculatus*, indicating that they had been secreted into the gut lumen [46]. Clearly, these enzymes fulfill a digestive function in these beetles.

According to our data, GVI1 and CMA3 are mannanases, which mainly break down galactomannan, and most likely also mannan. The ability of these enzymes to break down glucomannan as well is most likely due to their ability to cleave β-1,4-bonds of two adjoining mannose residues in the glucomannan backbone. The additional activity observed for CMA3 on carboxymethylcellulose may be explained by the ability of this enzyme to partially cleave β-1, 4-bonds between two glucose residues. This ability is most likely due to the fact that glucose is an isomer of mannose, and the high molecular similarity of both polysaccharides might lead to substrate recognition and subsequent cleavage by CMA3. However, we would like to point out that no such secondary enzymatic activity on carboxymethylcellulose was observed for GVI1. Thus, the ability of a GH5_10 mannanase to act as cellulase is the exception rather than the rule for this family of enzymes. In addition, we know that *G. viridula* encodes members of the GH45 family which are absent in *C. maculatus* [21]. Most GH45 proteins characterized from phytophagous beetles are endo-β-1,4-glucanases that can break down amorphous cellulose [22,25,26,47–49]. Thus, CMA3 may have evolved to break down amorphous cellulose on top of mannans to adapt to the loss of genes encoding GH45 cellulases present in other subfamilies of Chrysomelidae. Notably, bi-functional mannanase-cellulase enzymes such as CMA3 are rare among GH5 enzymes and were recorded only once for a GH5_1 protein in *Ruminococcus albus* (which is only able to cleave CMC and GluM as secondary activity to lichenan degradation) [50]. To our surprise, we discovered that CMA2 has lost its ability to degrade mannans and has evolved to break down xylan, another hemicellulosic polysaccharide. Xylanase activity in coleopteran species mediated by GH5s has so far been shown only

for two cerambycid beetles, but those cerambycid-derived enzymes belonged to a different GH5 subfamily than those investigated here [25,26]. Additionally, xylanases have been identified in the coffee berry borer, *Hypothenemus hampei* [51], and the mustard leaf beetle, *Phaedon cochleariae* [52], but those enzymes belonged to entirely different GH families (GH10 and GH11, respectively).

As cellulolytic and pectolytic enzymes are widely distributed among phytophagous beetles [21,24,25,53], implying they have an important biological function, it is striking that so few hemicellulolytic enzymes (such as xylanases and mannanases) have emerged in the course of Phytophaga beetle evolution. The distribution of hemicellulolytic enzymes in coleopteran species may hint at either a strong variation in the abundance of hemicellulosic polysaccharides in different plant species or plant organs, or at more specific biological requirements for breaking down hemicellulose according to the insect. We would also like to point out that it is still unclear whether symbiotic microbes are involved in plant cell wall degradation in beetles; if the microbes are, they may potentially be providing hemicellulolytic enzymes in the gut of insects that lack endogenous enzymes. The ability of *C. maculatus* to break down GalM, GluM, CMC and xylan by expressing only GH5_10 proteins likely evolved through subfunctionalization events. In an evolutionary context, this ability seems to be a consequential event, as no cellulases or xylanases of any other GH family were found in the larval transcriptome of *C. maculatus* [21]. Altogether, and taking into account our previous work on the *C. maculatus* GH28 proteins [23], this beetle species possesses the ability to almost completely break down the polysaccharides of the plant's primary cell wall by expressing only two GH families. CMA1, CMA4 and CMA5 (a likely allele of CMA4) exhibited no enzymatic activity against the substrates tested, although no substitution of important catalytic residues was observed. Interestingly, the expression of CMA4 transcripts in the gut tissue as well as the presence of the corresponding protein in the gut lumen of *C. maculatus* larvae have been reported elsewhere [45,46], suggesting that this protein plays an important role in the gut of *C. maculatus*. The inability to degrade any of the substrates we tested does not exclude the possibility that these proteins are still active enzymes for which no substrate has yet been found. This fact could also suggest a neo-functionalization of these proteins, but this possibility remains to be investigated.

To learn more about the physiological function of the *G. viridula* GH5_10 mannanase, we performed RNAi experiments. We successfully managed to knock down the expression of GVI1, which correlated with a reduction of enzymatic activity against galactomannan, indicating that GVI1 is the only mannanase expressed in the gut of *G. viridula*. However, we found no significant differences regarding either growth rate or mortality between GVI1-silenced larvae and the GFP control. These results would imply that GVI1 may not fulfill a primary digestive function, such as providing degraded plant cell wall polysaccharides (e.g. manno-oligomers or mannose) for metabolic purposes, e.g., glycolysis or fatty acid metabolism. We hypothesize instead that cleaving hemicellulosic components of the plant's cell wall is rather accessory and may facilitate exposure of plant cells to the insect, allowing *G. viridula* to gain access to and to benefit from simple sugars and proteins present in plant cells. Although the previous hypothesis may also elicit reduced growth and/or the increased mortality of silenced larvae, the activity of GVI1 may be compensated for by other glycoside hydrolase families encoded by *G.viridula*, such as GH45 cellulases and GH28 pectinases. But such speculation needs to be further examined by, for example, performing comparative expression profiling using RNA-Seq between silenced and control insects. Interestingly, Nogueira *et al*. demonstrated that *C. maculatus* treated with a cysteine peptidase inhibitor responded by up-regulating other digestive enzymes, including CMA3 [46]. This finding may suggest a compensation of inhibited proteases by other digestive enzymes or it is a stress response due to the decreasing

amount of nitrogen set free in the digestive tract of the animal. The former hypothesis seems rather unlikely, as mannan degradation does not directly increase nitrogen levels. Thus, we believe that a stress response is rather likely and this hypothesis has been already suggested elsewhere [54].

GH5 is a large multigene family, but its members are rarely found in Coleoptera or in insects in general; to date only three subfamilies (2, 8 and 10) have been identified [21,25–27,55]. Our investigation revealed that, according to data currently available, only 59 genes encoding GH5_10 are present within the tree of life, and, in fact, this gene family seems completely absent from plants and fungi. Of those 59 genes, only five are found in two species of Coleoptera. We can also exclude bacterial contamination because the GH5_10 genes of *G. viridula* and *C. maculatus* were found to harbor introns. Contamination by fungal-derived GH5s is also unlikely because, as previously mentioned, genes encoding GH5_10 seem absent in fungi. To our surprise, our phylogenetic analysis revealed that coleopteran-derived GH5_10 proteins did not cluster together with other insect counterparts but, rather, with those derived from three different species of mites belonging to the Oribatidae (Chelicerata). Additional GH5_10 sequences were identified in transcriptomes of insects belonging to orders other than Coleoptera, i.e. Zygentoma, Archaeognatha and Ephemeroptera. The latter sequences clustered together with collembolan- and crustacean-derived GH5_10 sequences. The patchy distribution of those proteins within arthropods indicates that the apparition of GH5_10 genes happened several times individually or else a massive gene loss has occurred in this phylum. As the latter hypothesis seems unlikely, we believe that the most parsimonious explanation for the appearance of GH5_10 genes in arthropods is the occurrence of several independent horizontal gene transfer (HGT) events, probably from bacteria to arthropod, occurring at several time points in the evolution of arthropods. However, we would like to note that because of the few bacteria-derived GH5_10 sequences currently available, our phylogenetic analysis has a poor resolution. Although we cannot fully support the hypothesis of an HGT event, the availability of an increasing amount of bacteria-derived GH5_10 sequences in the near future may solve this problem.

As we investigated the structure of the genes encoding GH5_10 in *G. viridula* and *C. maculatus*, we realized that the position and phase of the first and last introns are shared between these sequences. A logical explanation for this observation is that the most recent common ancestor of these two species of chrysomelid beetles possessed at least one GH5_10 gene harboring these two conserved introns. Subsequently, if the hypothesis for the acquisition of this gene family through an HGT event is true, this transfer should have happened at least in the most recent common ancestor of these two beetle species. Apart from *G. viridula* and *C. maculatus*, and according to the state of transcriptome and genome data currently available, genes encoding GH5_10 are apparently absent from other species of Chrysomelidae and even from species within the superfamily Chrysomeloidea and its sister superfamily Curculionoidea [21,25,28,53]. Altogether, it implies that a major gene loss happened among Chrysomelidae. An HGT event between these two species may also represent an alternative hypothesis, but this remains to be investigated.

Finally, the restriction of GH5_10 proteins to confined animal lineages may have represented an important factor to allow these animals to adapt to their food. Although the advantage of having mannanases is understandable for *C. maculatus* and other species of Bruchinae, it is less clear how *G. viridula* would benefit. In fact, species of Bruchinae often use the seeds of legumes as a food source. Legume seeds are notoriously rich in galactomannan, which is used as a storage sugar and subsequently as a source of energy during germination [7,56]. For an insect feeding on these seeds, the ability to break down galactomannan and use it as a potential source of energy represents a true advantage. We strongly believe that, within Bruchinae, the

presence of GH5_10 is common and not limited to *C. maculatus*. A similar situation has been observed in the coffee berry borer, *H. hampei*—a species that feeds on coffee beans, which are also very rich in mannans [7] -which has acquired a GH5_8 mannanase from bacteria through HGT [27]. On the other hand, our RNAi experiments indicated that knocking down the expression of GVI1, which correlated with a drastic reduction of the mannanase activity in the gut of silenced larvae, neither significantly decreased growth nor increased mortality compared to control larvae. We also could not find any information in the literature which would indicate that the content of mannans in *Rumex* sp. leaves is unusually high. Yet when GVI1 was put in contact with a preparation of plant cell wall isolated from *Rumex* leaves, mannan oligomers were visible on TLC, indicating that this enzyme likely contributes to breaking down the plant cell wall when the beetle feeds on its host plant.

Supporting information

S1 Fig. Amino acid alignment of beetle-derived GH5_10 proteins with two others for which the crystal structure has been resolved. Amino acid sequences were aligned without their predicted amino-terminal signal peptide. Conserved sites are depicted from dark to light blue, depending on the degree of amino acid identity. The two catalytic glutamate residues are indicated in red. Active site residues are indicated by arrows. The two reference sequences for which the crystal structure has been resolved are derived from the Antarctic springtail, *cryptopygus antarcticus* (CAN1, PDB: 4OOU_A), and from the blue mussel, *Mytilus edulis* (MED1, PDB: 2C0H_A).
(TIF)

S2 Fig. Tissue-specific expression of beetle-derived genes encoding GH5_10 proteins. Late-instar actively feeding larvae were dissected, and gut and rest bodies were used for total RNA preparation and quantitative RT-PCR. (A) The gene encoding GVI1 is significantly more expressed in the gut of *G. viridula* larvae compared to in the rest of the body. The gene expression is given as the copy number of GVI1 per 1000 molecules of RPS3 (control gene) ± SEM. (B) Genes encoding GH5_10 are significantly more expressed in the gut of *C. maculatus* larvae compared to in the rest of the body. The gene expression is given as the copy number of GVI1 per 1000 molecules of EF1α (control gene) ± SEM. Data were plotted using a log-transformed scale. Gene expression data were analyzed using paired t-tests (statistical values see S3 Table).
(TIF)

S3 Fig. Zymogram of the mannanase activity after RNAi in *G. viridula*. The same protein samples (day 4 post injection) as those described in Fig 4 were used for zymographic analyses. (A) 5 μg total proteins were loaded on a semi-native SDS-PAGE gel containing 0.1% (w/v) galactomannan. After the run, the gel was stained with Coomassie and used as a loading control. (B) 0.5 μg total proteins from the same samples were loaded on the same semi-native SDS-PAGE gel. After the run, this part of the gel was used to detect mannanase activity and activity bands were detected after staining with Congo red. iGH5: samples were prepared from *G. viridula* larvae injected with dsRNA targeting GVI1; iGFP: samples were prepared from larvae injected with dsRNA targeting GFP and used as controls; NIC: non-injected control larvae.
(TIF)

S4 Fig. Action of GVI1 on a preparation of plant cell wall from *Rumex obtusifolius* leaves. GVI1 was heterologously expressed in *Sf*9 cells and crude enzyme extract was incubated with a preparation of protein-free plant cell wall (PCW) isolated from *R. obtusifolius* leaves. Results were analyzed on TLC. A reaction in which GVI1 had been omitted was included as a control.

In addition, the PCW was also incubated with a commercially available control cellulase preparation (CCP) isolated from *Trichoderma reesei*. Several standards were used: from mannose (M1) to mannohexaose (M6); from glucose (G1) to cellopentaose (G5); from xylose (X1) to xylotriose (X3).
(TIF)

S5 Fig. Conservation of intron position and phase between *G. viridula* and *C. maculatus* GH5_10 genes. The amino acid sequences of *G. viridula* GVI1 and *C. maculatus* CMA1 to CMA4 were aligned using MUSCLE. The sequence corresponding to the signal peptide is indicated in bold. The *G. viridula* GVI1 gene was amplified by PCR using gDNA as a template. The sequences corresponding to the *C. maculatus* GH5_10 genes were retrieved from a genome draft assembly of this species (http://www.beanbeetles.org/genome/). Missing sequence data for the *C. maculatus* GH5_10 genes are indicated in gray. Intron positions and phase are indicated by colored amino acids. Amino acids in green correspond to the insertion of a phase 0 intron. Amino acids in red correspond to the insertion of a phase 1 intron. Amino acids in blue correspond to the insertion of a phase 2 intron.
(TIF)

S1 Table. List of primers used in this study.
(PDF)

S2 Table. Details of the amino acid sequences used for the phylogenetic analysis.
(PDF)

S3 Table. Statistical analysis of tissue specific gene expression (S2 Fig).
(PDF)

Acknowledgments

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch for technical support. We thank Emily Wheeler, Boston, for editorial assistance. We are also thankful to Roy Kirsch for his input on experimental design and for fruitful discussions.

Author Contributions

Conceptualization: André Busch, Yannick Pauchet.

Data curation: André Busch, Yannick Pauchet.

Formal analysis: André Busch, Grit Kunert, Yannick Pauchet.

Funding acquisition: David G. Heckel.

Investigation: André Busch.

Methodology: André Busch, Grit Kunert, Yannick Pauchet.

Resources: David G. Heckel.

Supervision: Yannick Pauchet.

Validation: André Busch, Grit Kunert, Yannick Pauchet.

Writing – original draft: André Busch, Yannick Pauchet.

Writing – review & editing: André Busch, Grit Kunert, David G. Heckel, Yannick Pauchet.

References

1. Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol 6: 850–861. https://doi.org/10.1038/nrm1746 PMID: 16261190
2. Cosgrove DJ (2014) Re-constructing our models of cellulose and primary cell wall assembly. Curr Opin Plant Biol 22: 122–131. https://doi.org/10.1016/j.pbi.2014.11.001 PMID: 25460077
3. Rodriguez-Gacio Mdel C, Iglesias-Fernandez R, Carbonero P, Matilla AJ (2012) Softening-up mannan-rich cell walls. J Exp Bot 63: 3976–3988. https://doi.org/10.1093/jxb/ers096 PMID: 22553284
4. Meier H, Reid JSG (1982) Reserve Polysaccharides Other Than Starch in Higher Plants. In: Loewus FA, Tanner W, editors. Plant Carbohydrates I. Berlin Heidelberg New York: Springer-Verlag. pp. 418–471.
5. Buckeridge MS, dos Santos HP, Tine MAS (2000) Mobilisation of storage cell wall polysaccharides in seeds. Plant Physiol Biochem 38: 141–156.
6. Matheson NK (1990) Mannose-based Polysaccharides. In: Dey PM, editor. Methods in Plant Biochemistry. London: Academic Press Limited. pp. 371–413.
7. Buckeridge MS (2010) Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. Plant Physiol 154: 1017–1023. https://doi.org/10.1104/pp.110.158642 PMID: 20855518
8. Soni H, Kango N (2013) Microbial Mannanases: Properties and Applications. In: Shukla P, Pletschke BI, editors. Advances in Enzyme Biotechnology. New Dehli: Springer India. pp. 41–56.
9. Malgas S, van Dyk JS, Pletschke BI (2015) A review of the enzymatic hydrolysis of mannans and synergistic interactions between beta-mannanase, beta-mannosidase and alpha-galactosidase. World J Microbiol Biotechnol 31: 1167–1175. https://doi.org/10.1007/s11274-015-1878-2 PMID: 26026279
10. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37: D233–D238. https://doi.org/10.1093/nar/gkn663 PMID: 18838391
11. Larsson AM, Anderson L, Xu B, Munoz IG, Uson I, Janson JC, et al. (2006) Three-dimensional crystal structure and enzymic characterization of beta-mannanase Man5A from blue mussel *Mytilus edulis*. J Mol Biol 357: 1500–1510. https://doi.org/10.1016/j.jmb.2006.01.044 PMID: 16487541
12. Xu B, Hagglund P, Stalbrand H, Janson JC (2002) endo-beta-1,4-Mannanases from blue mussel, *Mytilus edulis*: purification, characterization, and mode of action. J Biotechnol 92: 267–277. PMID: 11689251
13. Ootsuka S, Saga N, Suzuki K, Inoue A, Ojima T (2006) Isolation and cloning of an endo-beta-1,4-mannanase from Pacific abalone *Haliotis discus hannai*. J Biotechnol 125: 269–280. https://doi.org/10.1016/j.jbiotec.2006.03.008 PMID: 16621092
14. Zahura UA, Rahman MM, Inoue A, Tanaka H, Ojima T (2011) cDNA cloning and bacterial expression of an endo-beta-1,4-mannanase, AkMan, from *Aplysia kurodai*. Comp Biochem Physiol B Biochem Mol Biol 159: 227–235. https://doi.org/10.1016/j.cbpb.2011.05.001 PMID: 21601647
15. King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ, et al. (2010) Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. Proc Natl Acad Sci U S A 107: 5345–5350. https://doi.org/10.1073/pnas.0914228107 PMID: 20212162
16. Song JM, Nam KW, Kang SG, Kim CG, Kwon ST, Lee YH (2008) Molecular cloning and characterization of a novel cold-active beta-1,4-D-mannanase from the Antarctic springtail, *Cryptopygus antarcticus*. Comp Biochem Physiol B Biochem Mol Biol 151: 32–40. https://doi.org/10.1016/j.cbpb.2008.05.005 PMID: 18579426
17. Aspeborg H, Coutinho PM, Wang Y, Brumer H 3rd, Henrissat B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol 12: 186. https://doi.org/10.1186/1471-2148-12-186 PMID: 22992189
18. Angelov A, Liebl S, Ballschmiter M, Bomeke M, Lehmann R, Liesegang H, et al. (2010) Genome sequence of the polysaccharide-degrading, thermophilic anaerobe *Spirochaeta thermophila* DSM 6192. J Bacteriol 192: 6492–6493. https://doi.org/10.1128/JB.01023-10 PMID: 20935097
19. Weiner RM, Taylor LE 2nd, Henrissat B, Hauser L, Land M, Coutinho PM, et al. (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2–40 T. PLoS Genet 4: e1000087. https://doi.org/10.1371/journal.pgen.1000087 PMID: 18516288
20. Marvaldi AE, Duckett CN, Kjer KM, Gillespie JJ (2009) Structural alignment of 18S and 28S rDNA sequences provides insights into phylogeny of Phytophaga (Coleoptera: Curculionoidea and Chrysomeloidea). Zool Scr 38: 63–77.

21. Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH (2010) Diversity of beetle genes encoding novel plant cell wall degrading enzymes. PLoS One 5: e15635. https://doi.org/10.1371/journal.pone.0015635 PMID: 21179425

22. Girard C, Jouanin L (1999) Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae*. Insect Biochem Mol Biol 29: 1129–1142. PMID: 10612046

23. Kirsch R, Gramzow L, Theissen G, Siegfried BD, Ffrench-Constant RH, Heckel DG, et al. (2014) Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. Insect Biochem Mol Biol 52: 33–50. https://doi.org/10.1016/j.ibmb.2014.06.008 PMID: 24978610

24. Kirsch R, Wielsch N, Vogel H, Svatos A, Heckel DG, Pauchet Y (2012) Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. BMC Genomics 13: 587. https://doi.org/10.1186/1471-2164-13-587 PMID: 23116131

25. McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, et al. (2016) Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. Genome Biol 17: 227. https://doi.org/10.1186/s13059-016-1088-8 PMID: 27832824

26. Pauchet Y, Kirsch R, Giraud S, Vogel H, Heckel DG (2014) Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle *Apriona japonica*. Insect Biochem Mol Biol 49: 1–13. https://doi.org/10.1016/j.ibmb.2014.03.004 PMID: 24657889

27. Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, et al. (2012) Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proc Natl Acad Sci U S A 109: 4197–4202. https://doi.org/10.1073/pnas.1121190109 PMID: 22371593

28. Vega FE, Brown SM, Chen H, Shen E, Nair MB, Ceja-Navarro JA, et al. (2015) Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer, *Hypothenemus hampei*. Sci Rep 5: 12525. https://doi.org/10.1038/srep12525 PMID: 26228545

29. Pauchet Y, Saski CA, Feltus FA, Luyten I, Quesneville H, Heckel DG (2014) Studying the organization of genes encoding plant cell wall degrading enzymes in *Chrysomela tremula* provides insights into a leaf beetle genome. Insect Mol Biol 23: 286–300. https://doi.org/10.1111/imb.12081 PMID: 24456018

30. Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E (2006) Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls. Plant Methods 2: 10. https://doi.org/10.1186/1746-4811-2-10 PMID: 16729891

31. Iseli C, Ambrosini G, Bucher P, Jongeneel CV (2007) Indexing strategies for rapid searches of short words in genome sequences. PLoS One 2: e579. https://doi.org/10.1371/journal.pone.0000579 PMID: 17593978

32. R-Core-Team (2014) R: a language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria.

33. Pinheiro J, Bates D, DebRoy S, Sarkar D, R core Team (2016) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–128.

34. Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM (2009) Mixed Effects Models and Extensions in Ecology with R: Springer-Verlag New York.

35. Crawley MJ (2012) The R Book: Wiley-Blackwell.

36. Kapustin Y, Souvorov A, Tatusova T, Lipman D (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. Biol Direct 3: 20. https://doi.org/10.1186/1745-6150-3-20 PMID: 18495041

37. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42: D490–495. https://doi.org/10.1093/nar/gkt1178 PMID: 24270786

38. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465–W469. https://doi.org/10.1093/nar/gkn180 PMID: 18424797

39. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30: 2725–2729. https://doi.org/10.1093/molbev/mst197 PMID: 24132122

40. Dodd D, Moon YH, Swaminathan K, Mackie RI, Cann IK (2010) Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. J Biol Chem 285: 30261–30273. https://doi.org/10.1074/jbc.M110.141788 PMID: 20622018

41. Foong FC, Doi RH (1992) Characterization and comparison of *Clostridium cellulovorans* endoglucanases-xylanases EngB and EngD hyperexpressed in *Escherichia coli*. J Bacteriol 174: 1403–1409. PMID: 1735727

42. Kim MK, An YJ, Song JM, Jeong CS, Kang MH, Kwon KK, et al. (2014) Structure-based investigation into the functional roles of the extended loop and substrate-recognition sites in an endo-beta-1,4-d-mannanase from the Antarctic springtail, *Cryptopygus antarcticus*. Proteins 82: 3217–3223. https://doi.org/10.1002/prot.24655 PMID: 25082572

43. Sinha RN (1959) The Hydrogen-Ion Concentration in the Alimentary Canal of Beetles Infesting Stored Grain and Grain Products. Ann Entomol Soc Am 52: 763–765.

44. Scheller HV, Ulvskov P (2010) Hemicelluloses. Annu Rev Plant Biol 61: 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315 PMID: 20192742

45. Chi YH, Salzman RA, Balfe S, Ahn JE, Sun W, Moon J, et al. (2009) Cowpea bruchid midgut transcriptome response to a soybean cystatin—costs and benefits of counter-defence. Insect Mol Biol 18: 97–110. https://doi.org/10.1111/j.1365-2583.2008.00854.x PMID: 19196350

46. Nogueira FC, Silva CP, Alexandre D, Samuels RI, Soares EL, Aragao FJ, et al. (2012) Global proteome changes in larvae of *Callosobruchus maculatus* Coleoptera:Chrysomelidae:Bruchinae) following ingestion of a cysteine proteinase inhibitor. Proteomics 12: 2704–2715. https://doi.org/10.1002/pmic.201200039 PMID: 22833537

47. Lee SJ, Kim SR, Yoon HJ, Kim I, Lee KS, Je YH, et al. (2004) cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. Comp Biochem Physiol B Biochem Mol Biol 139: 107–116. https://doi.org/10.1016/j.cbpc.2004.06.015 PMID: 15364293

48. Lee SJ, Lee KS, Kim SR, Gui ZZ, Kim YS, Yoon HJ, et al. (2005) A novel cellulase gene from the mulberry longicorn beetle, *Apriona germari*: Gene structure, expression, and enzymatic activity. Comp Biochem Physiol B Biochem Mol Biol 140: 551–560. https://doi.org/10.1016/j.cbpc.2004.12.003 PMID: 15763510

49. 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: 260–267. https://doi.org/10.1016/j.gene.2012.10.046 PMID: 23137634

50. Iakiviak M, Mackie RI, Cann IK (2011) Functional analyses of multiple lichenin-degrading enzymes from the rumen bacterium *Ruminococcus albus* 8. Appl Environ Microbiol 77: 7541–7550. https://doi.org/10.1128/AEM.06088-11 PMID: 21890664

51. Padilla-Hurtado B, Florez-Ramos C, Aguilera-Galvez C, Medina-Olaya J, Ramirez-Sanjuan A, Rubio-Gomez J, et al. (2012) Cloning and expression of an endo-1,4-beta-xylanase from the coffee berry borer, *Hypothenemus hampei*. BMC Res Notes 5: 23. https://doi.org/10.1186/1756-0500-5-23 PMID: 22233686

52. Pauchet Y, Heckel DG (2013) The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer. Proc Biol Sci 280: 20131021. https://doi.org/10.1098/rspb.2013.1021 PMID: 23698014

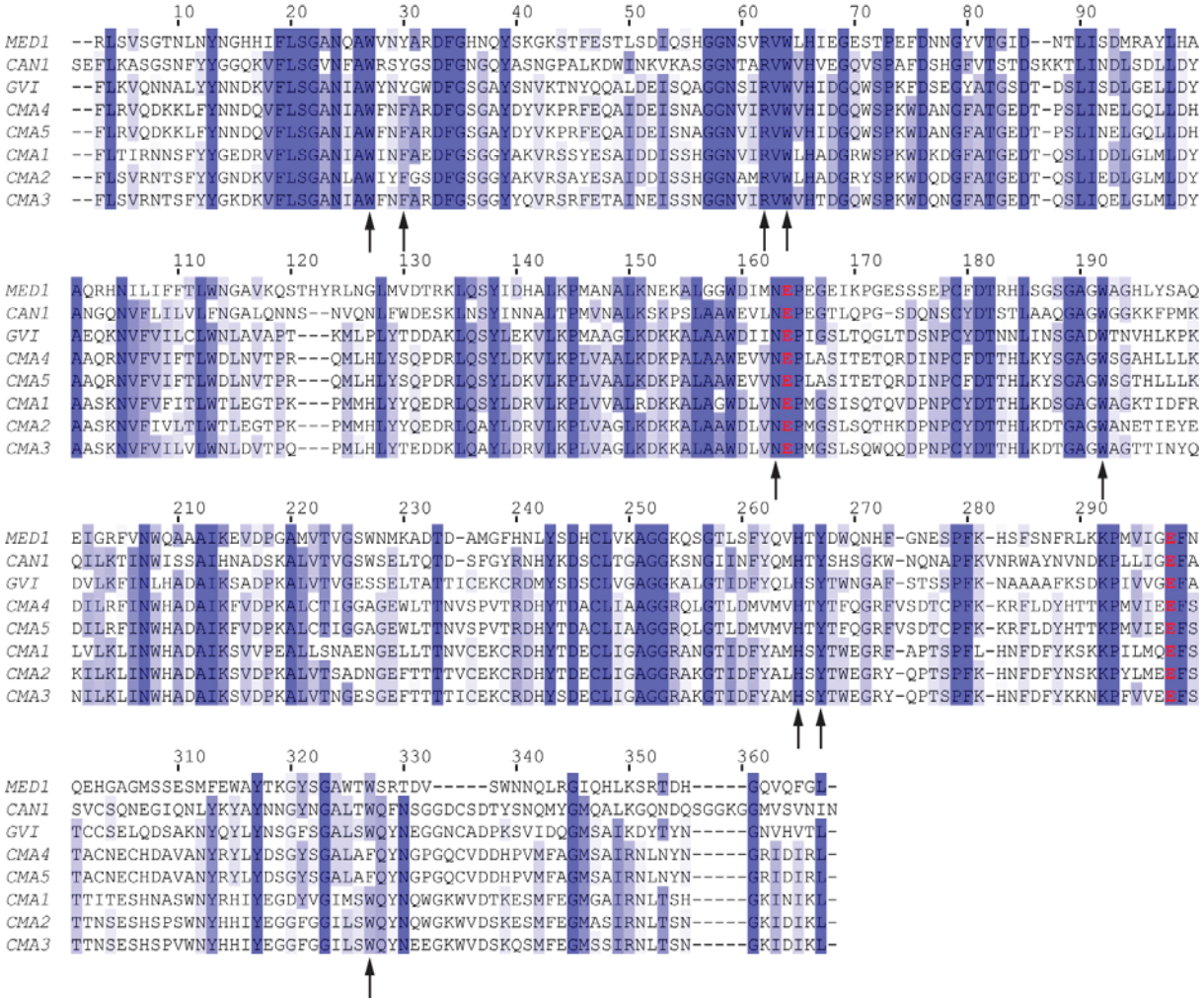
53. Keeling CI, Yuen MM, Liao NY, Docking TR, Chan SK, Taylor GA, et al. (2013) Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol 14: R27. https://doi.org/10.1186/gb-2013-14-3-r27 PMID: 23537049

54. Kuwar SS, Pauchet Y, Vogel H, Heckel DG (2015) Adaptive regulation of digestive serine proteases in the larval midgut of *Helicoverpa armigera* in response to a plant protease inhibitor. Insect biochemmol biol 59: 18–29.

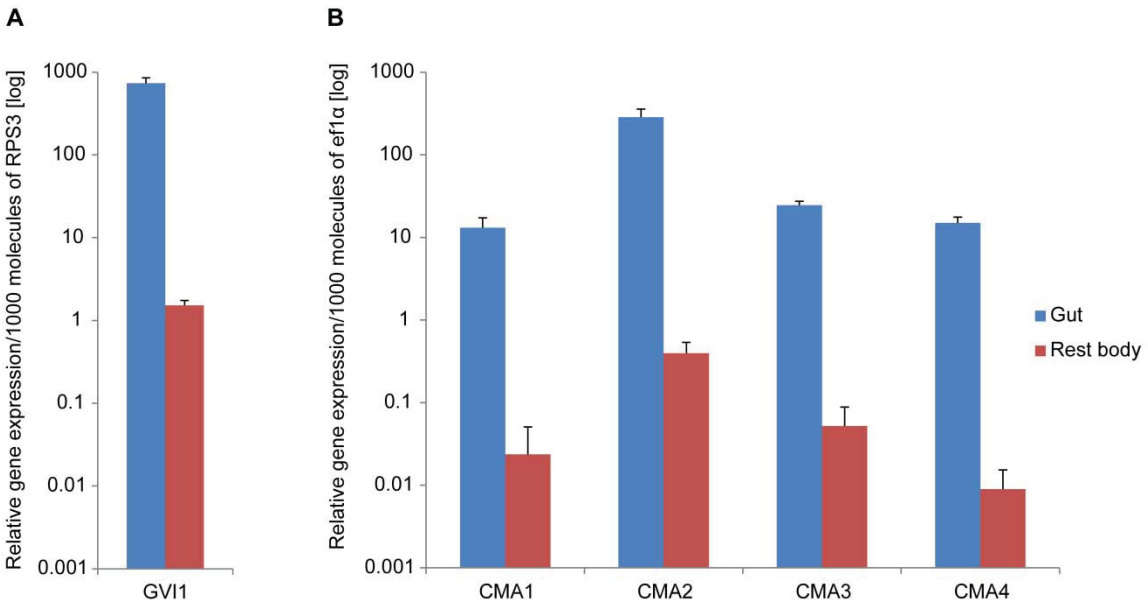
55. Wei YD, Lee KS, Gui ZZ, Yoon HJ, Kim I, Zhang GZ, et al. (2006) Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, *Apriona germari*. Comp Biochem Physiol B Biochem Mol Biol 145: 220–229. https://doi.org/10.1016/j.cbpb.2006.07.007 PMID: 16945565

56. Scherbukhin VD, Anulov OV (1999) Legume seed galactomannans (review). Appl Biochem Microbiol 35: 229–244.

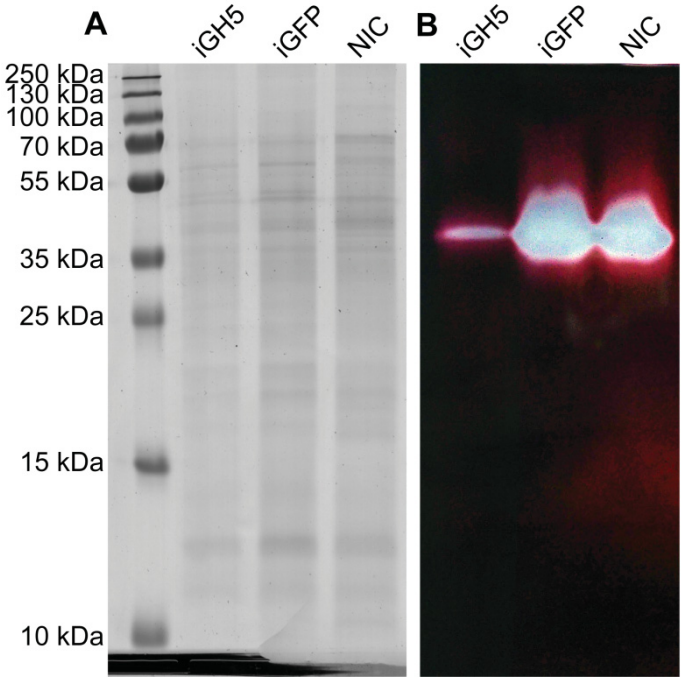
Supplementary Material



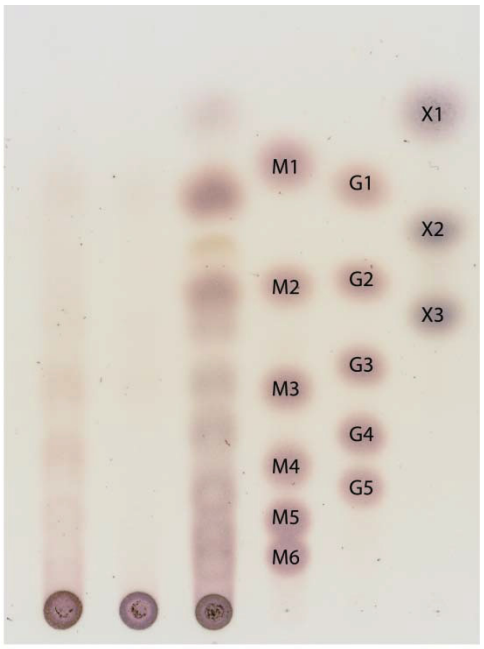
S1 Fig. Amino acid alignment of beetle-derived GH5_10 proteins with two others for which the crystal structure has been resolved. Amino acid sequences were aligned without their predicted amino-terminal signal peptide. Conserved sites are depicted from dark to light blue, depending on the degree of amino acid identity. The two catalytic glutamate residues are indicated in red. Active site residues are indicated by arrows. The two reference sequences for which the crystal structure has been resolved are derived from the Antarctic springtail, *Cryptopygus antarcticus* (CAN1, PDB: 4OOU_A), and from the blue mussel, *Mytilus edulis* (MED1, PDB: 2C0H_A).



S2 Fig. Tissue-specific expression of beetle-derived genes encoding GH5_10 proteins. Late-instar actively feeding larvae were dissected, and gut and rest bodies were used for total RNA preparation and quantitative RT-PCR. (A) The gene encoding GVI1 is significantly more expressed in the gut of *G. viridula* larvae compared to in the rest of the body. The gene expression is given as the copy number of GVI1 per 1000 molecules of RPS3 (control gene) ± SEM. (B) Genes encoding GH5_10 are significantly more expressed in the gut of *C. maculatus* larvae compared to in the rest of the body. The gene expression is given as the copy number of GVI1 per 1000 molecules of EF1α (control gene) ± SEM. Data were plotted using a log-transformed scale. Gene expression data were analyzed using paired t-tests (statistical values see S3 Table).

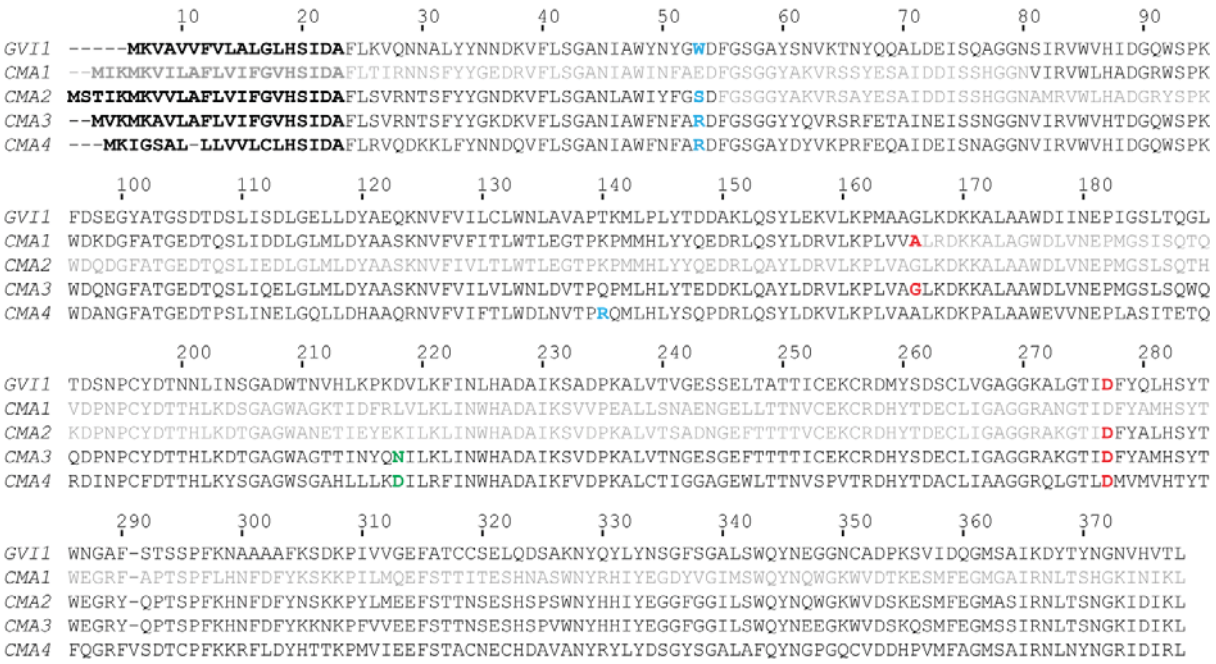


S3 Fig. Zymogram of the mannanase activity after RNAi in *G. viridula*. The same protein samples (day 4 post injection) as those described in Fig 4 were used for zymographic analyses. (A) 5 µg total proteins were loaded on a semi-native SDS-PAGE gel containing 0.1% (w/v) galactomannan. After the run, the gel was stained with Coomassie and used as a loading control. (B) 0.5 µg total proteins from the same samples were loaded on the same semi-native SDS-PAGE gel. After the run, this part of the gel was used to detect mannanase activity and activity bands were detected after staining with Congo red. iGH5: samples were prepared from *G. viridula* larvae injected with dsRNA targeting GVI1; iGFP: samples were prepared from larvae injected with dsRNA targeting GFP and used as controls; NIC: non-injected control larvae.



GVI1	+	-	-	Standards
PCW	+	+	+	
CCP	-	-	+	

S4 Fig. Action of GVI1 on a preparation of plant cell wall from *Rumex obtusifolius* leaves. GVI1 was heterologously expressed in *Sf*9 cells and crude enzyme extract was incubated with a preparation of protein-free plant cell wall (PCW) isolated from *R*. *obtusifolius* leaves. Results were analyzed on TLC. A reaction in which GVI1 had been omitted was included as a control. In addition, the PCW was also incubated with a commercially available control cellulase preparation (CCP) isolated from *Trichoderma reesei*. Several standards were used: from mannose (M1) to mannohexaose (M6); from glucose (G1) to cellopentaose (G5); from xylose (X1) to xylotriose (X3).



S5 Fig. Conservation of intron position and phase between *G. viridula* and *C. maculatus* GH5_10 genes. The amino acid sequences of *G*. *viridula* GVI1 and *C*. *maculatus* CMA1 to CMA4 were aligned using MUSCLE. The sequence corresponding to the signal peptide is indicated in bold. The *G*. *viridula* GVI1 gene was amplified by PCR using gDNA as a template. The sequences corresponding to the *C*. *maculatus* GH5_10 genes were retrieved from a genome draft assembly of this species (http://www.beanbeetles.org/genome/). Missing sequence data for the *C*. *maculatus* GH5_10 genes are indicated in gray. Intron positions and phase are indicated by colored amino acids. Amino acids in green correspond to the insertion of a phase 0 intron. Amino acids in red correspond to the insertion of a phase 1 intron. Amino acids in blue correspond to the insertion of a phase 2 intron.

S1 Table. List of primers used in this study.

Gene	Primer name	Sequence (5' – 3')	Function
GVI1	GVI1_ORF_F	ACCATGGAAGTCGCTGTGGTATTCG	Cloning pIB/V5-His TOPO/TA
GVI1	GVI1_ORF_R	AAGGGTCACATGGACATTGCCGTTG	Cloning pIB/V5-His TOPO/TA
CMA1	CMA1_ORF_F	ACCATGGTCAAGATGAAGGTGAT	Cloning pIB/V5-His TOPO/TA
CMA1	CMA1_ORF_R	CAACTTTATATTAATCTTTCCATGG	Cloning pIB/V5-His TOPO/TA
CMA2	CMA2_ORF_F	ACCATGGCCACGATCAAGATGAAGGT	Cloning pIB/V5-His TOPO/TA
CMA2	CMA2_ORF_R	CAACTTTATATCAATCTTTCCGTTG	Cloning pIB/V5-His TOPO/TA
CMA3	CMA3_ORF_F	ACCATGGTCAAGATGAAAGCGGT	Cloning pIB/V5-His TOPO/TA
CMA3	CMA3_ORF_R	CAACTTTATGTCAATCTTTCCATTG	Cloning pIB/V5-His TOPO/TA
CMA4	CMA4_ORF_F	ACCATGGAAATTGGATCCGCTCTGCT	Cloning pIB/V5-His TOPO/TA
CMA4	CMA4_ORF_R	CAGCCGAATATCAATCCTCCCA	Cloning pIB/V5-His TOPO/TA
CMA5	CMA5_ORF_F	ACCATGGAACTTGGATCCGCACTGCT	Cloning pIB/V5-His TOPO/TA
CMA5	CMA5_ORF_R	CAGCCTAATATCAATCCTCCCA	Cloning pIB/V5-His TOPO/TA
GVI1	GVI1_dsRNA_F	TAATACGACTCACTATAGGGAGCACCAACAACTTGATCAACAG	dsRNA generation
GVI1	GVI1_dsRNA_R	TAATACGACTCACTATAGGGAGTTCTTGAAGGGACTGGAGGT	dsRNA generation
GFP	GFP_dsRNA_F	TAATACGACTCACTATAGGCACATGAAGCAGCACGACTT	dsRNA generation
GFP	GFP_dsRNA_R	TAATACGACTCACTATAGGTGCTCAGGTAGTGGTTGTCG	dsRNA generation
GVI1	GVI1_F	CAACTATGGATGGGACTTCG	Quantitative PCR
GVI1	GVI1_R	GATGTGCACCCAAACTCTGA	Quantitative PCR
GVI_RpS3	GVI_RpS3_F	GCAGGATCCGTGAGTTGAC	Quantitative PCR
GVI_RpS3	GVI_RpS3_R	ACAGGCCTCGATTGGCTAC	Quantitative PCR
CMA1	CMA1_F	AGAAGACACCCAATCGCTGA	Quantitative PCR
CMA1	CMA1_R	CTGCAGCCTATCTTCCTGGT	Quantitative PCR
CMA2	CMA2_F	ACTCTGTGGACTTTGGAGGG	Quantitative PCR
CMA2	CMA2_R	TTGTCTTTCAGGCCTGCAAC	Quantitative PCR
CMA3	CMA3_F	TCAGGAGAGTTCACCACCAC	Quantitative PCR
CMA3	CMA3_R	CTGTGCATGGCGTAGAAGTC	Quantitative PCR
CMA4	CMA4_F	AACCCTTGCTTTGACACCAC	Quantitative PCR
CMA4	CMA4_R	CACAGCGCTTTAGGATCCAC	Quantitative PCR
CMA_EF1α	CMA_EF1α_F	AAGGCCTCCACACACATAGG	Quantitative PCR
CMA_EF1α	CMA_EF1α_R	AAGGTTGATCGTCGTTCTGG	Quantitative PCR

S2_Table. Details of the amino acid sequences used for the phylogenetic analysis.

Specie name	abbreviation	Accession number	Notes
Bacteria			
Fibrobacter succinogenes subsp. succinogenes S85	FSU1	ACX74338.1	Fibrobacteres
Flammeovirga sp. MY04	FLA1	ANQ49303.1	Bacteroidetes
Flammeovirga sp. MY04	FLA2	ANQ52741.1	Bacteroidetes
Flammeovirga yaeyamensis	FYA1	ACA05117.2	Bacteroidetes
Melioribacter roseus P3M-2	MRO1	AFN74927.1	Ignavibacteriae
Saccharophagus degradans 2-40	SDE1	ABD81545.1	Proteobacteria (gamma)
Spirochaeta thermophila DSM 6192	STH1	ADN01068.1	Spirochaetes
Spirochaeta thermophila DSM 6578	STH2	AEJ60366.1	Spirochaetes
Uncultured bacterium	UBA1	AEV59732.1	
Vibrio natriegens	VNA1	ANQ24656.1	Proteobacteria (gamma)
Insects			
Gastrophysa viridula	GVI1	ADU33333.1	Arthropoda (Coleoptera)
Callosobruchus maculatus	CMA1	ADU33271.1	Arthropoda (Coleoptera)
Callosobruchus maculatus	CMA2	ADU33272.1	Arthropoda (Coleoptera)
Callosobruchus maculatus	CMA3	ADU33273.1	Arthropoda (Coleoptera)
Callosobruchus maculatus	CMA4	ADU33274.1	Arthropoda (Coleoptera)
Tricholepidion gertschi	TGE1	GASO02036642	Arthropoda (Zygentoma)
Tricholepidion gertschi	TGE2	GASO02042472	Arthropoda (Zygentoma)
Thermobia domestica	TDO1	GASN02045897	Arthropoda (Zygentoma)
Thermobia domestica	TDO2	GASN02045380	Arthropoda (Zygentoma)
Machilis hrabei	MHR1	GAUM02040243	Arthropoda (Archaeognatha)
Eurylophella sp. AD-2013	EUR1	GAZG02013172	Arthropoda (Ephemeroptera)
Isonychia bicolor	IBI1	GAXA02038640	Arthropoda (Ephemeroptera)
Collembola			
Cryptopygus antarcticus	CAN1	ABV68808.1	Arthropoda
Orchesella cincta OC12957	OCI1	GAMM01012947	Arthropoda
Sminthurus viridis	SVI1	GATZ02024785	Arthropoda
Folsomia candida	FCA1	GASX02002882	Arthropoda
Pogonognathellus sp. AD-2013	POG1	GATD02008031	Arthropoda
Tetrodontophora bielanensis	TBI1	GAXI02022226	Arthropoda
Crustacea			
Limnoria quadripunctata	LQU1	ADE58567.1	Arthropoda
Limnoria quadripunctata	LQU2	ADE58568.1	Arthropoda
Limnoria quadripunctata	LQU3	ADE58569.1	Arthropoda
Hyalella azteca	HAZ1	XP_018022370.1	Arthropoda
Daphnia pulex	DPU1	EFX71596.1	Arthropoda
Daphnia pulex	DPU2	EFX71597.1	Arthropoda
Proasellus grafi	PGR1	HAEX01030896	Arthropoda
Talitrus saltator	TSA1	GDUJ01040948	Arthropoda
Cherax quadricarinatus	CQU1	HACK01027957	Arthropoda
Chelicerata			
Nothrus palustris	NPA1	GEYJ01076133	Arthropoda
Nothrus palustris	NPA2	GEYJ01112851	Arthropoda
Nothrus palustris	NPA3	GEYJ01073735	Arthropoda
Platynothrus peltifer	PPE1	GEYZ01027687	Arthropoda
Platynothrus peltifer	PPE2	GEYZ01016327	Arthropoda
Platynothrus peltifer	PPE3	GEYZ01010130	Arthropoda
Steganacarus magnus	SMA1	GEYQ01012964	Arthropoda
Steganacarus magnus	SMA2	GEYQ01047325	Arthropoda
Gastropoda			
Aplysia kurodai	AKU1	BAJ60954.1	Mollusca
Biomphalaria glabrata	BGL1	AAV91523.1	Mollusca
Haliotis discus discus	HDI1	ACJ12612.1	Mollusca
Haliotis discus discus	HDI2	ACJ12613.1	Mollusca
Haliotis discus hannai	HDI3	BAE78456.1	Mollusca
Limacina antarctica	LAN1	GDRM01030628	Mollusca
Elysia timida	ETI1	GBRM01102742	Mollusca
Deroceras reticulatum	DRE1	JW037389	Mollusca
Pomacea canaliculata	PCA1	GBZZ01045211	Mollusca

S3 Table. Statistical analysis of tissue specific gene expression (Fig S2).

Species	Gene	t-value	p-value
G.viridula	GH5	16.092	0.004
C. maculatus	GH5-1	4.525	0.006
	GH5-2	9.019	0.003
	GH5-3	30.468	<0.001
	GH5-4	7.422	0.002

3.2. Manuscripts II

Cellulose degradation in *Gastrophysa viridula* (Coleoptera: Chrysomelidae): functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity

André Busch[1], Grit Kunert[2], Natalie Wielsch[3] and Yannick Pauchet[1]

[1]Department of Entomology, [2]Department of Biochemistry and [3]Research Group Mass Spectrometry, Max Planck Institute for Chemical Ecology, Jena, Germany

Manuscript published in
Insect Molecular Biology

Doi: 10.1111/imb.12500
License number: 4481880557424

Abstract

Cellulose is a major component of the primary and secondary cell walls in plants. Cellulose is considered to be the most abundant biopolymer on Earth and represents a large potential source of metabolic energy. Yet, cellulose degradation is rare and mostly restricted to cellulolytic microorganisms. Recently, various metazoans, including leaf beetles, have been found to encode their own cellulases, giving them the ability to degrade cellulose independently of cellulolytic symbionts. Here, we analyzed the cellulosic capacity of the leaf beetle *Gastrophysa viridula,* which typically feeds on *Rumex* plants. We identified three putative cellulases member of two glycoside hydrolase (GH) families, namely GH45 and GH9. Using heterologous expression and functional assays, we demonstrated that both GH45 proteins are active enzymes, in contrast to the GH9 protein. One GH45 protein acted on amorphous cellulose as an endo-β-1,4-glucanase, whereas the other evolved to become an endo-β-1,4-xyloglucanase. We successfully knocked down the expression of both GH45 genes using RNAi, but no changes in weight gain or mortality were observed compared to control insects. Our data indicated that the breakdown of these polysaccharides in *G. viridula* may facilitate access to plant cell content, which is rich in nitrogen and simple sugars.

Keywords

Chrysomelidae; cellulase; xyloglucanase; GH9; GH45; RNAi

Introduction

As the sun's electromagnetic energy is gathered by plants and converted into chemical energy, most of it is stored as glucose-derived polysaccharides. Among these are polysaccharides contributing to the architecture of the plant cell wall, such as the hemicelluloses (e.g. xyloglucan and mannans) or pectin (e.g. polygalacturonic acid). However, the most abundant polysaccharide is believed to be cellulose (Bayer et al. 1998), which consists of β-1,4-linked anhydroglucose residues forming an unbranched polysaccharide chain. After biogenesis at the cellulose synthase complex, the individual chains attach to each other by hydrogen bonds and yield a (para)crystalline structure (Chang 1981). Yet, there are "weak points" on the surface of the cellulose polymer that are not crystallized and these are referred to as amorphous regions (Ruel et al. 2012). Although it is assumed that cellulose crystallinity is low in primary cell walls (Cosgrove 2014), we still lack a complete understanding of how the native cellulose is packed with regards to the proportion of amorphous and crystalline cellulose (Knox 2008; Saxena 2007).

Nonetheless, cellulose theoretically represents an abundant source of energy to any organism which has evolved the ability to degrade it. However, in nature, the degradation of cellulose seems to be difficult, as it is recalcitrant to hydrolysis (Bayer et al. 1998). Thus, only a few organisms have evolved the ability to degrade cellulose enzymatically, and, for a long time, cellulose degradation was believed to be restricted to the domain of microorganisms, such as saprophytic fungi (Schulein 1997), plant pathogenic parasites (Chambost J.P. 1987; Py et al. 1991), or mutualistic symbionts in the gut of insects (Breznak and Brune 1994) or in the rumen of ruminant animals (Rincon et al. 2001). In 1998, the first metazoan-derived cellulase was discovered (Watanabe et al. 1998), and since then, cellulases and other plant cell wall degrading enzymes (PCWDEs) have been found continuously in multicellular organisms though their distribution is patchy. Those discoveries ultimately changed the dogmatic view that the cellulolytic system in metazoans is dependent solely on microbes. Today, the presence of endogenous cellulolytic enzymes in multicellular organisms spans the phyla Nematoda, Arthropoda and Mollusca (Girard and Jouanin

1999; Kikuchi et al. 2004; Pauchet et al. 2014a; Sakamoto and Toyohara 2009; Shelomi et al. 2014; Smant et al. 1998; Willis et al. 2011; Xu et al. 2001).

To degrade cellulosic material, cellulases have evolved two types of activity and are, accordingly, named either exo-β-1,4-glucanases (EC 3.2.1.91) or endo-β-1,4-glucanases (EC 3.2.1.4). Exoglucanases (cellobiohydrolases) attack the polysaccharide from its non-reducing terminal regions, releasing cellobiose and occasionally cellotriose (Takahashi et al. 2010). Cellobiohydrolases are able to break down amorphous as well as crystalline cellulose. In contrast, endoglucanases are able to attack only amorphous regions, releasing randomly sized fragments of cellulose. Both types of cellulases act synergistically in order to efficiently break down the cellulose network (Kostylev and Wilson 2012).

According to amino acid sequence similarities, cellulases are classified into distinct families of glycoside hydrolases (GHs) (Henrissat and Bairoch 1993). In insects, the most commonly distributed endogenous cellulolytic families are GH5, GH9 and GH45 (Fischer et al. 2013). The first indication of a coleopteran-derived GH45 involved in cellulose degradation was described in the leaf beetle *Phaedon cochleariae* (Coleoptera: Chrysomelidae) in 1999 (Girard and Jouanin 1999). But not until 2004 was the first recombinant GH45 characterized as an active cellulase within the beetles (Lee et al. 2004). In the following years, several other beetle-derived GH45s were identified and functionally characterized (Chang et al. 2012; Lee et al. 2004; Lee et al. 2005; McKenna et al. 2016; Mei et al. 2015; Valencia et al. 2013; Valencia Jimenez et al. 2014; Xia et al. 2013) or found in transcriptome as well as proteome data (Busconi et al. 2014; Kirsch et al. 2012; Pauchet et al. 2010). Today, the superfamilies Chrysomeloidea (leaf beetles and longhorned beetles) and Curculionoidea (weevils), generally referred to as the Phytophaga clade (Marvaldi et al. 2009), are commonly found to encode proteins of the GH45 family (Pauchet et al. 2010). Interestingly, to date and to the best of our knowledge, there are only two cases of endogenous GH9s known to be present in beetles: first, a cellulolytically active GH9 described in the tenebrionid *Tribolium castaneum* (Willis et al. 2011) and

second, a GH9 from the cerambycid *Anoplophora glabripennis*, which did not show enzymatic activity under the experimental conditions tested (McKenna et al. 2016).

Here, we analyzed the ability of the green dock beetle *Gastrophysa viridula* (Coleoptera: Chrysomelidae) to break down various polysaccharides typically associated with the plant cell wall. Our main objective was to identify the enzymes responsible for the breakdown of cellulose in this species and to estimate their biological relevance. Previously, we analyzed the larval gut transcriptome generated from *G. viridula* and identified a single transcript encoding a putative cellulase, a member of the GH45 family, which we named GH45-1 (Pauchet et al. 2010). Re-analysis of this transcriptome allowed us to identify a transcript encoding a second putative cellulase member of the GH9 family. We demonstrate here that neither GH45-1 nor GH9 possesses the ability to break down any of the cellulosic substrates we tested. Using protein purification, we identified a protein responsible for the breakdown of amorphous cellulose in the gut of *G. viridula* larvae. This protein happened to be a second GH45 (GH45-2) that was previously not covered by the transcriptome analysis. After functionally expressing the corresponding proteins, we demonstrate that GH45-1 has evolved to become an endo-β-1,4-xyloglucanase, and our demonstration represents the first case of such activity within the GH45s; GH45-2, in contrast and as expected for this GH family, is a endo-β-1,4-glucanase acting on amorphous cellulose. Finally, in order to assess the biological importance of these GH45s in *G. viridula*, we used RNAi to knock down the expression of the corresponding genes. Although successful, the knockdown of the expression of the two GH45 genes induced no striking phenotype (weight gain and mortality) compared to control insects.

Results

Enzymes present in the *G. viridula* gut can attack a range of plant cell wall polysaccharides

We used dialyzed and desalted gut content in an in-tube assay against various kinds of celluloses, hemicelluloses and pectins, and the breakdown products were visualized by TLC (Fig. 1B, D, E). We were able to show enzymatic activity against glucomannan and galactomannan (Fig. 1B), which, in both cases, are degraded down to the monomer as well as to larger oligomers. In addition, we found that the gut content is able to degrade various pectins (Fig. 1C). In particular, polygalacturonic acid is broken down to mainly monomer, dimer as well as larger oligomers, whereas citrus pectin is mostly degraded into trimer and larger oligomers with traces of monomer and dimer. Pectin derived from apple is degraded to monomer, trimer and larger oligomers, but lacks dimers. The patterns of esterified pectin and beet pectin show the least degradation, most likely due to substitutions of the pectin polymer with either methyl ester (esterified pectin) or acetyl groups (beet pectin). These results are in agreement with our previous work confirming both mannanase (Busch et al. 2017) and pectinase activity (Kirsch et al. 2014) in *G. viridula*. More important, we found that each cellulosic substrate tested (carboxymethyl cellulose (CMC), cellobiose, cellotriose, cellopentaose, microcrystalline cellulose and filter paper) was degraded down to glucose (Fig. 1A). For carboxymethyl cellulose (CMC), cello-oligomers also appear as “smears” on TLC. We do note, however, that microcrystalline cellulose (Avicel) shows only a faint signal, matching a monomer in size and indicating that cellulolytic proteins of the gut content exhibit only restricted activity towards microcrystalline cellulose. Additionally, we found that xyloglucan was highly degraded. Xyloglucan breakdown products correlate with glucose, cellobiose and cellotriose (as well as with larger oligomers), but, based on the fact that glucose moieties within the xyloglucan backbone can be highly substituted with xylose (which, in turn, may be substituted with galactose and/or fucose), the observed breakdown products can reflect any combination of those polysaccharides. This experiment clearly demonstrated the ability of *G. viridula* to effectively degrade most of the major polysaccharides present in the plant cell wall.

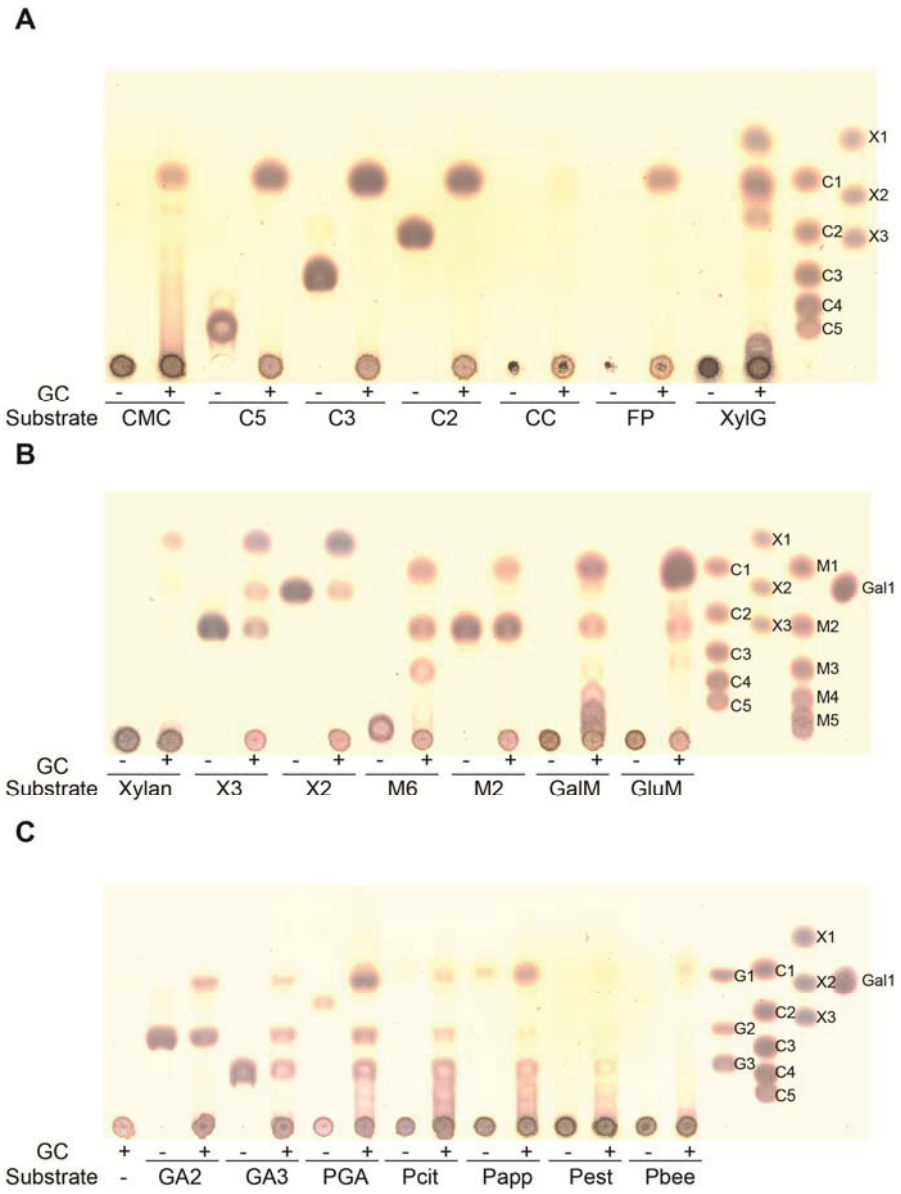


Fig. 1. Enzymatic activity of *G. viridula* gut content against several plant-cell-wall-derived polysaccharides. (A,B,C) TLC of *G. viridula* gut content incubated with several plant-cell-wall-derived polysaccharides for 16 h at 40 °C and pH 5.0. Breakdown products were visualized using 0.2 % orcinol in methanol/sulphoric acid. C1-C5 = cellulose-oligomers; CMC = carboxymethyl cellulose; CC = crystalline cellulose; FP = filter paper; XylG = xyloglucan; Xyl1-Xyl3 = xylan-oligomers; Man1-Man6 = mannose-oligomers; GluM = glucomannan; GalM = galactomannan; Gal1 = galactose; GA1-GA3 = galacturonic acid-oligomers; PGA = polygalacturonic acid; Pcit = citrus pectin; Papp = apple pectin; Pest = esterified pectin; Pbee = beet pectin.

Investigation of candidate genes encoding putative cellulases

Our previous transcriptome analysis revealed a transcript encoding a putative cellulase of the glycoside hydrolase family 45 (GH45-1; accession: ADU33334.1). The full-length amplicon of GH45-1 was cloned into a pIB-V5/His TOPO vector in frame with a V5/His_6-Tag and was subsequently used for transient transfection in insect *Sf*9 cells. According to the presence of an amino-terminal signal peptide in the sequence of GH45-1, the target recombinant protein was secreted into the culture medium and its successful expression was validated by Western blot using Anti-V5 Antibody. GH45-1 was successfully expressed with an apparent molecular weight of ~35 kDa (Fig. 2A). Although the protein's theoretically expected size was ~26 kDa, we reasoned that the increased size of the target protein was due to the artificially added V5/His_6-tag as well as to post-translational modifications such as glycosylation. To analyze the cellulolytic properties of the recombinant GH45-1, we tested the crude protein extract on an agarose-diffusion plate containing 0.1 % CMC (Fig. 2B). To our surprise, no cellulase activity was detected. We re-screened the transcriptome of *G. viridula* and discovered a partial sequence with strong similarities to proteins of the glycoside hydrolase family 9 (GH9). As GH9s are well known for their cellulolytic properties (Kim et al. 2008; Watanabe et al. 1998; Willis et al. 2011), we decided to RACE-amplify the GH9 and we obtained a full-length sequence (accession: MG875329). We then cloned and heterologously expressed the *G. viridula* GH9 and tested it for its cellulolytic ability. We validated the successful expression by Western blot, yielding a protein signal with an apparent molecular weight of ~60 kDa (Fig. 2A and B), which is in accordance with its estimated size (~53.4 kDa). Once more, no activity against CMC could be observed. The confirmed activity of the gut lumen in response to several cellulosic substrates indicated that either the *G. viridula* transcriptome was incomplete or a gene(s) coding for protein families of yet unknown cellulolytic ability was (were) present. To track the cellulolytic enzyme(s) present in the larval gut of *G. viridula*, we performed a protein purification scheme.

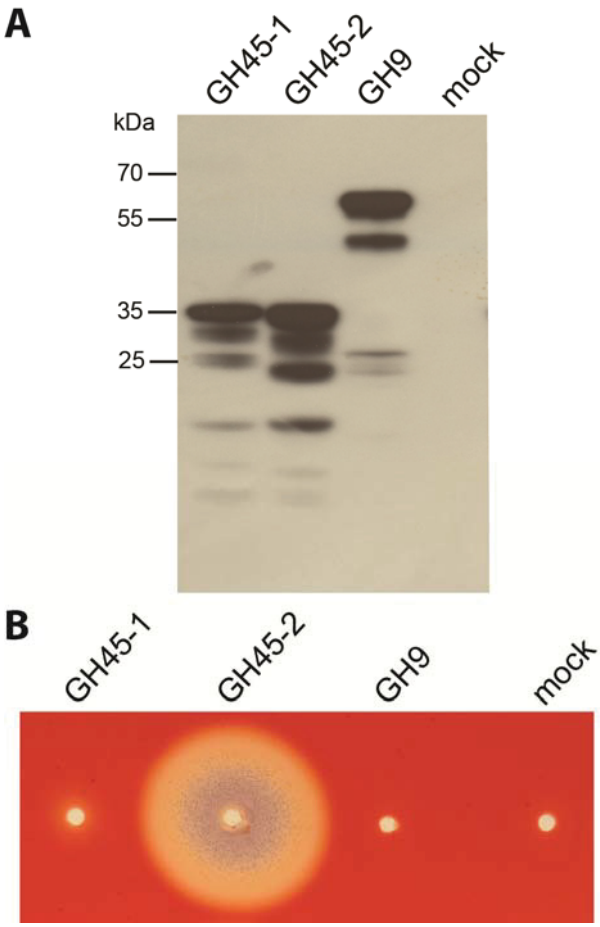


Fig. 2. Western Blot and enzymatic activity assays of target recombinant GH45 and GH9 proteins. (A) Western blot of target recombinant proteins expressed in frame with a V5/His(6)-Epitope in insect *Sf*9 cells. Detection was carried out by chemiluminescence using an anti-V5-HRP antibody. Molecular weight markers are indicated to the left of the Western blot in kilodalton (kDa). (B) Agarose-diffusion-assay containing 0.1 % CMC. The assay was incubated for 16 h at 40 °C and pH 5.0 with target recombinant proteins, and activity halos were revealed by Congo red staining.

Purification of a cellulolytically active GH45 protein from the larval gut

We used 100 *G. viridula* larval guts to perform three consecutive protein purification steps. Using agarose-diffusion-assays containing 0.1 % CMC, we traced CMCase activity within each purification step. To identify potential proteins acting as cellulases, we performed a LC-MS/MS analysis based on the final cellulolytically active fraction of the purified gut content (Fig. S1). The resulting data were searched against the NCBInr database using Mascot. In addition, the peptides were interpreted *de novo*

and searched against the same database as well as the subdatabase ‘Insecta’ of NCBInr using MS-BLAST. We identified two putative cellulase proteins: the first belonged to the glycoside hydrolase family 48 (GH48) and was identified both by Mascot and MS-BLAST searches and corresponded to a *G. viridula* GH48 (Table S2); the second identified protein had peptides matching GH45 proteins from another chrysomelid beetle, *Phaedon cochleariae*, but no matches with *G. viridula* GH45-1 in our Mascot and MS-BLAST searches. Indeed, no protein fragment belonging to GH45-1 appeared at all in the investigated samples. Altogether, our purification scheme uncovered a novel GH45 in *G. viridula* with strong similarities to GH45 of other closely related leaf beetles which was not identified during our first transcriptome analysis (Pauchet et al. 2010).

In summary, both GH45 and GH48 family members discovered by LC-MS/MS are known to have cellulolytic abilities in microbes and are likely cellulase candidates in *G. viridula*. Until now, however, GH48s in beetles have not been confirmed to be cellulases and are indeed believed to have lost their cellulolytic abilities (Sukharnikov et al. 2012). To date, they are solely known for their distinct ability to degrade chitin (Fujita et al. 2006). In contrast, GH45s have been frequently characterized as cellulases in beetles (Chang et al. 2012; Girard and Jouanin 1999; Lee et al. 2004; Pauchet et al. 2014a). Based on this knowledge, we focused our attention on the novel GH45 rather than the GH48.

In order to clone the cDNA corresponding to the newly identified GH45 protein, we first designed degenerated primers (Table S1) based on sequence information from various GH45s found in closely related leaf beetle species. As a PCR template, we used RACE-ready cDNAs generated from larval gut tissue and performed gradient PCRs. The resulting PCR fragments of appropriate size were cloned and sequenced. Using BLAST homology searches, we were able to identify a partial sequence matching a GH45 derived from *P. cochleariae* instead of GH45-1 derived from *G. viridula*. Based on that fragment, we designed specific primers and were able to successfully amplify a full-length sequence using RACE-PCR. Using BLAST searches, we also confirmed that the newly cloned *G. viridula* GH45 protein was absent from the larval gut transcriptome. Thus, from here on the novel GH45 will be

referred to as GH45-2 (accession: MG875330). Finally, the peptides obtained by LC-MS/MS from the partially purified GH45 matched 100 % to the deduced protein sequence of GH45-2.

G. viridula GH45 proteins harbor distinct enzymatic properties

To test the cellulolytic property of GH45-2, the ORF was cloned and heterologously expressed in *Sf*9 cells. Positive expression, confirmed by Western blot, yielded an apparent molecular weight of ~35 kDa (Fig. 2A). The culture medium containing the heterologously expressed GH45-2 was deposited on an agarose diffusion plate containing 0.1 % CMC; an activity halo was observed on the plate after staining (Fig. 2B), indicating that GH45-2 was an active endo-β-1,4-glucanase.

To further analyze the enzymatic characteristics of beetle-derived GH45s and GH9, we performed in-tube assays with a battery of plant-cell-wall-derived poly- and oligosaccharides. The breakdown products were visualized on TLC (Fig. 3). GH45-2 was active on CMC and regenerated amorphous cellulose (RAC) by releasing dimers, trimers and tetramers, confirming that GH45-2 was indeed an endo-β-1,4-glucanase. We then tested several cellulose oligomers against GH45-2, ranging from cellotriose to cellohexaose. Only cellopentaose and cellohexaose were degraded by the enzyme, releasing dimers, trimers and tetramers. Noteworthy was the appearance of a tetramer after cellopentaose degradation. Although a tetramer implies exo-activity of GH45-2, no corresponding monomer was detected, suggesting that GH45-2 may have the ability to catalyze a transglycosylation reaction between two dimers, as has been described before in other glycoside hydrolase families (Tanabe et al. 2003). Unexpectedly, we discovered that GH45-1 possessed the ability to break down xyloglucan instead of cellulose. The breakdown products correlated with heptamers and octamers, but may differ due to highly substituted glucose moieties within the xyloglucan polysaccharide (as described above), indicating that GH45-1 was an endo-β-1,4-xyloglucanase (EC 3.2.1.151). According to the literature, GH45s are known solely for their activity towards cellulose (Guo et al. 2008; Kikuchi et al. 2004; Liu et al. 2010; Mcgavin and Forsberg 1988; Pauchet et al. 2014a). In conclusion, we believe we have discovered the first case of a GH45 with specific xyloglucanase activity.

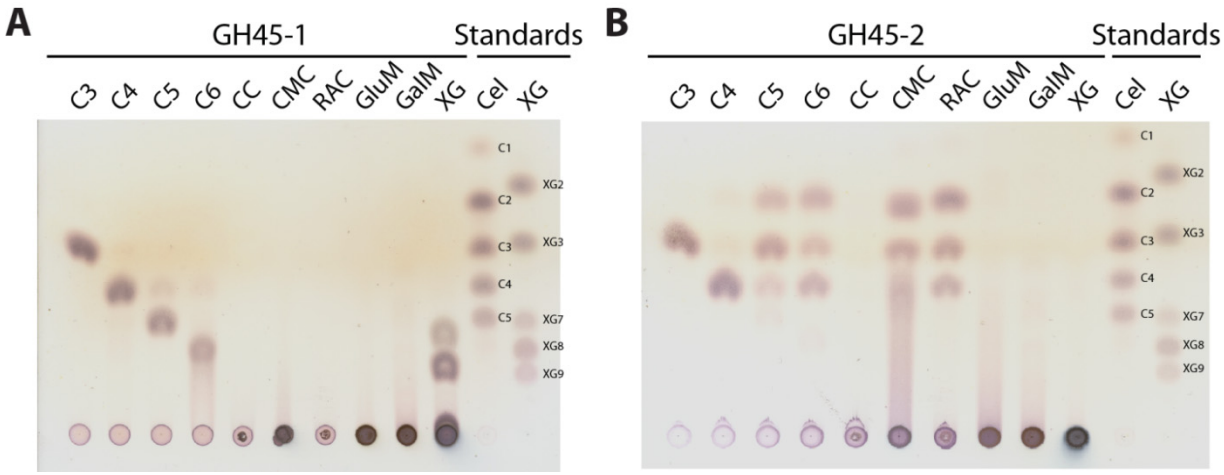


Fig. 3. Enzymatic breakdown products of various plant-cell-wall-derived polysaccharides tested against recombinant GH45 proteins. Incubation of (A) GH45-1 and (B) GH45-2 with several plant-cell-wall-derived polysaccharides for 16 h at 40 °C and pH 5.0. Breakdown products were analyzed on TLC and visualized using 0.2 % orcinol in methane/sulphoric acid (9:1) under continuous heating. CMC = carboxymethyl cellulose; CC = crystalline cellulose (Avicel); RAC = regenerated amorphous cellulose; GluM = glucomannan; GalM = galactomannan; XG = xyloglucan; standards: C1-C5 = glucose to cellopentaose; XG2 - XG9 = isoprimeverose to xyloglucan nona saccharide.

In addition, we also attempted to characterize the GH9 protein with the same enzymatic assays as described above. Surprisingly, no activity was detected on any of the substrates tested, indicating either that the *G. viridula* GH9 protein has lost its enzymatic ability or that we were unable to find its substrate.

pH and temperature optima of the *G. viridula* GH45 proteins

The enzymatic performance of GH45-1 was monitored using xyloglucan as substrate. The optimal pH of GH45-1 laid was about pH 5.0 (Fig. S2A) with an optimal temperature around 40 °C (Fig. S2B). We then tested GH45-2 using CMC as a substrate. Similar to our findings regarding GH45-1, GH45-2 performed best under acidic conditions, however, with a much broader pH spectra, ranging from pH 2.0 to 6.0 (Fig. S2A), and with an optimal temperature around 50 °C (Fig. S2B). We also investigated the thermal stability of both GH45s by incubating them for 16 h starting at their optimal temperatures and increasing to 80 °C without substrate. Then, the residual activity was measured at their optimal temperature (Fig. S2C). Our data

showed that both enzymes were highly stable between 40 and 60 °C with minor (GH45-1) or no loss (GH45-2) of activity. In summary, each enzyme performed best under acidic conditions, a finding which correlated well with gut pH measurements from closely related beetle species (Sinha 1959a).

Tissue-specific gene expression and gene expression knockdown of GH45s in *G.viridula*

By performing qPCR comparing the midgut tissue and the corresponding rest of the body of *G. viridula* larvae (Fig. S3), we showed that transcripts encoding both GH45s were significantly more expressed in the gut compared to in the rest of the body. In contrast, the expression of the gene encoding the GH9 protein was extremely low in both tissues. This finding reinforced the fact that both GH45s have a digestive function, whereas the GH9 has likely lost digestive significance in *G. viridula*.

We then investigated the biological relevance of both GH45 proteins by using gene expression knockdown (i45-1 and i45-2 treatments) by means of RNAi (Fig. 4A and B). Gene expression analyses were carried out four and eight days after dsRNA injection as well as at the adult stage (seven days post-eclosion). We were able to significantly down-regulate the expression of the genes encoding GH45-1 (likelihood ratio = 72.514, P <0.001) and GH45-2 (likelihood ratio = 52.715, P < 0.001). More precisely, we were able to reduce the gene expression of GH45-1 to 96.9 % on day four and 89.6 % on day eight post-injection as well as 52.1 % in adults compared to a GFP control injection (iGFP). Similarly, the gene expression of GH45-2 went down to 89.3 % on day four, 83.3 % on day eight and 95.8 % at the adult stage compared to iGFP (detailed statistical values see supplementary table S4). In summary, our RNAi treatment was successful in knocking down the expression of the genes encoding both GH45 proteins; this knockdown of expression remained stable until adulthood.

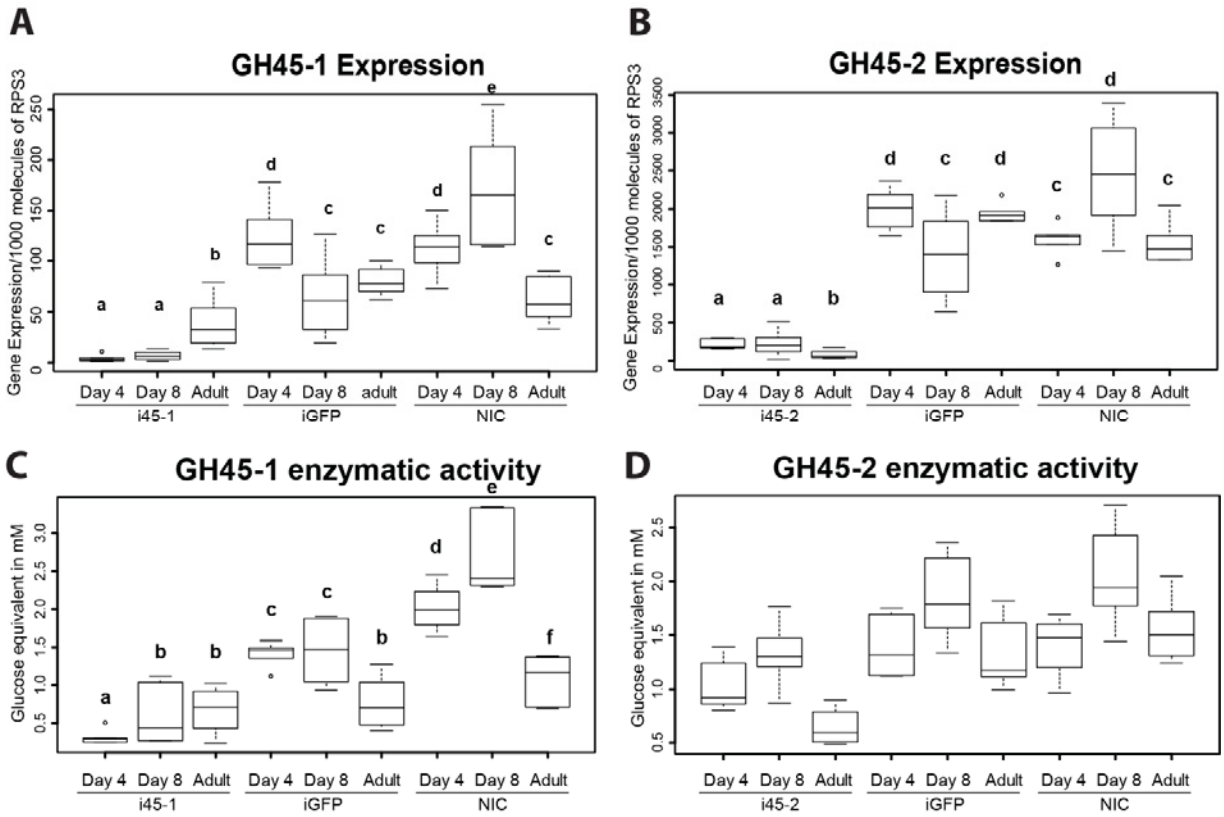


Fig. 4. Knockdown of the expression of the genes encoding GH45 proteins using RNA interference. Late first-instar larvae were injected with double-stranded RNA (dsRNA) targeting GH45-1 (i45-1) and -2 (i45-2) or GFP (iGFP) as controls. A non-injected control (NIC) was also included. Larvae were sampled on days four and eight post-injection. Adults were sampled one week after eclosion. Groups of three insects (six replicates per group) were snap-frozen in liquid nitrogen before being ground into a fine powder. Half of the powder was used for total RNA preparation and subsequent qPCR, and the other half was used for protein extraction and enzyme assays. (A, B) expression of the respective GH45 genes expressed per 1000 molecules of the reference gene after dsRNA treatment at different time points post-injection, and (C, D) corresponding enzymatic activity depicted as amounts of reducing sugar equivalents enzymatically released. Different letters indicate significant differences between treatments over time ($P < 0.05$). If changes over time were not different between treatments (D), letters are not applicable and were omitted. For details on the statistics, please refer to the Materials and Methods section and supplementary table S4.

To test whether the knockdown of the expression of the genes encoding both GH45 proteins affected the abundance of the corresponding proteins, we performed

enzymatic activity assays (Fig. 4C and D, statistical values supplemental table S4). The enzymatic activity of treatment i45-1 and i45-2 was investigated and compared to iGFP on day four, day eight and as adults. Larvae of the i45-1 treatment exhibited significantly reduced enzymatic activity against xyloglucan (likelihood ratio = 43.640, $P < 0.001$). In detail, the enzymatic activity against xyloglucan was reduced in larvae of the i45-1 treatment down to 84.3 % on day four and 58.9 % on day eight post-injection as well as 12.0 % at the adult stage compared to insects of the iGFP control treatment. Taken together, the reduction of xyloglucanase activity in insects of the i45-1 treatment was significant compared to the level of activity in control insects and correlated well with the knockdown of the expression of the gene encoding GH45-1.

A more complex situation applied to i45-2 treatments: cellulase activity was significantly different between the treatments (F= 15.989, $P < 0.001$). The activity in i45-2 treated insects was significantly lower than in both control groups. In general, the cellulase activity changed significantly over time (F= 14.548, $P < 0.001$), with the highest activity on day eight post-injection. These changes over time did not differ between treatments (F=0.809, $P = 0.526$). Interestingly, the resulting enzymatic down-regulation in the i45-2 treatment did not match the corresponding gene expression pattern. The resulting enzymatic down-regulation reached 26.2 % on day four and 25.0 % on day eight post-injection as well as 50.9 % in adults compared to iGFP control insects, amounts that are clearly inconsistent with the corresponding i45-2 gene expression pattern (knockdown >83 % at all time points). This discrepancy indicates that either protein levels of GH45-2 in i45-2 treatments are less affected by the knockdown of the corresponding gene, than what we observed after knocking down the expression of GH45-1, or more than one cellulase is present in the larval gut.

To examine why the reduction of enzymatic activity against cellulose and xyloglucan did not reach the same levels as the corresponding gene knockdown suggested, we conducted a zymogram analysis using either CMC or xyloglucan as a substrate (Fig. S4). We compared both i45-1 and i45-2 treatments to non-injected (NIC) and iGFP control insects on day four post-injection. We observed a single activity band

appearing in both controls on zymogram containing xyloglucan. No signal for the i45-1 treatment was observed, confirming (i) that only one endo-active xyloglucanase is present in *G. viridula* and (ii) that, although the encoding gene is down-regulated, residual xyloglucanase activity due to GH45-1 likely accounts for the discrepancy observed between gene knockdown and corresponding enzymatic activity. Analyzing the zymogram containing CMC, we observed activity bands corresponding to several endo-β-1,4-glucanase isozymes in each treatment, indicating that GH45-2 may likely not be the only endo-acting cellulase present in *G. viridula*. We observed only a slight reduction of the intensity of a single isozyme (the one harboring the highest apparent molecular weight) in the i45-2 treatment compared to both controls, this isozyme likely corresponded to GH45-2. Although we were able to down-regulate the expression of the gene encoding GH45-2, it is clear that the genome of *G. viridula* encodes additional endo-active cellulases and that these greatly influence cellulose breakdown.

To get insight into whether gene silencing of the target GH45s affected weight gain and mortality, we observed larvae for seven consecutive days post-injection (Fig. 5) and analyzed them on the last day of larval development (day eight post-injection). Although we expected to see a decrease in weight gain after dsRNA treatment, we were unable to detect a significant reduction among insects in the i45-1 treatment ($F = 0.891$; $P = 0.431$) or the i45-2 treatment ($F = 1.151$, $P = 0.344$). Similarly, we did not observe any significant differences in mortality among insects in either the i45-1 ($\chi^2 = 1.0976$, $P = 0.5777$) or the i45-2 ($\chi^2 = 0.42353$, $P = 0.8092$) treatments. In summary, although the knockdown of the genes encoding GH45-1 and GH45-2 turned out to be successful, a corresponding decrease in enzymatic activity could be observed only among insects in the i45-1 treatment. In addition, we were unable to find an influence of the knockdown of both GH45-encoding genes on any of the life history traits we analyzed. Altogether, our results suggest a secondary role for GH45 enzymes rather than a primary function in digestion.

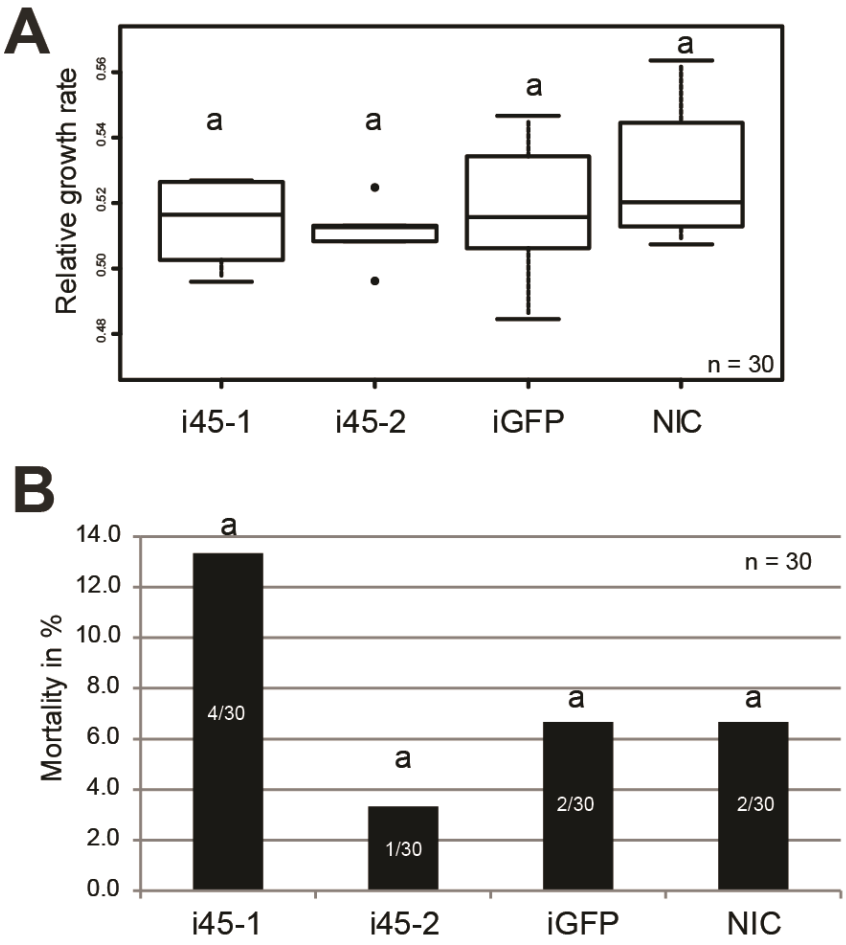


Fig. 5. Effect of GH45 gene silencing on growth rate and mortality. Late first-instar larvae were injected with double-stranded RNA (dsRNA) targeting GH45-1 (i45-1) and -2 (i45-2) or GFP (iGFP) as controls. A non-injected control (NIC) was also included. (A) Groups of five insects (six replicates per treatment) were observed on days one and eight post-injection and growth rates were calculated. A one-way ANOVA statistical test was applied to the data. (B) The number of dead larvae per treatment was recorded during the eight days of the experiment. Mortality data were analyzed using the equality of proportions test. Different letters indicate significant differences ($P < 0.05$); n = 30.

Significance for *G. viridula's* ability to express a xyloglucanase and a endo-acting cellulase

Considering that the primary plant cell wall consists of up to 30 % of cellulose and equally high amounts of xyloglucan (20-25 %) (Fry 1989; Hayashi 1989; Schultink et al. 2014), we asked whether GH45-1 and GH45-2 can break down xyloglucan and

cellulose under more natural conditions (Fig. S5). To test this hypothesis, we extracted plant primary cell wall from *Rumex obtusifolius* leaves and incubated it with either GH45-1 or GH45-2. The breakdown products were then analyzed on TLC. We found that GH45-1 was indeed able to release oligomers from the primary cell wall compared to untreated plant cell wall extract. This ability indicated the presence of plant-cell-wall-derived polysaccharides -- likely xyloglucan -- which can be accessed and digested by GH45-1. In contrast, no breakdown products appeared for plant cell wall treated with GH45-2, indicating that cellulose might be harder to access and/or has other chemical properties, such as a crystalline configuration, under more natural conditions.

Discussion

Our previous analysis of the larval gut transcriptome of *G. viridula* revealed the presence of at least one putative cellulase, a GH45 protein (Pauchet et al. 2010). We demonstrated here that this GH45 protein (GH45-1) was inactive on all the cellulosic substrates we tested. After rescreening the *G. viridula* transcriptome, we discovered another gene encoding another putative cellulase, this time, a GH9 protein that we had originally overlooked. Again, we showed here that the recombinant GH9 protein was inactive on all the cellulosic substrates tested. The ability of the digestive system of *G. viridula* to break down cellulose was obvious (Fig. 1); thus, we hypothesized that yet another cellulase, either one not covered by the transcriptome analysis or one belonging to a yet-unknown protein family, was present in the larval gut of this beetle. Indeed, after a three-step purification process of larval gut proteins, we enriched the samples with proteins able to break down CMC and eventually identified a second GH45 (GH45-2). When heterologously expressed, GH45-2 acted as an endo-β-1,4-glucanase. The ability of *G. viridula* GH45-2 to break down amorphous cellulose in an endo-acting manner was in accordance with other, previously described, GH45 cellulases in insects, nematodes or microbes (Girard and Jouanin 1999; Kikuchi et al. 2004; Mcgavin and Forsberg 1988). To our surprise, we found that GH45-1 had lost its assumed ability to degrade cellulose but had evolved the ability to degrade xyloglucan, the major hemicellulose in the plant primary cell wall (Pauly et al. 2013). The breakdown products generated by *G. viridula* GH45-1 also suggested that this enzyme was an endo-β-1,4-xyloglucanase. To our knowledge, this is the first report of a GH45 protein exhibiting activity against xyloglucan. In phytophagous beetles, only one other protein has been described as able to break down xyloglucan, namely a member of the subfamily 2 of GH5 (GH5_2) from the cerambycid *Anoplophora glabripennis* (McKenna et al. 2016). However, because this particular GH5 had acquired a bifunctional mode of action, it was also able to act on amorphous cellulose. The ability to break down xyloglucan has also been discovered in the Phasmatodea or stick insects. Here, an orthologous clade of GH9 proteins had evolved the bi-functional ability to break down xyloglucan in addition to cellulose (Shelomi et al. 2016). Interestingly, xyloglucanase activity in insects seems to arise

only within glycoside hydrolase families which are primarily known for their ability to degrade cellulose, i.e. GH5_2, GH9 and now GH45, suggesting that cellulolytic enzymes of these families may be evolutionary adapted to substrate shifts.
The ability of a protein to shift from using cellulose to using xyloglucan as a substrate can be explained by the similarities of the structure of these two polysaccharides. The molecular structure of cellulose consists of an unbranched succession of β-1,4-linked glucose residues. Xyloglucan shares the same backbone structure as cellulose but has additional α-1,6-linked xylose substitutions on two out of four glycosyl residues; these substitutions are in turn occasionally decorated with galactose and/or fucose residues (Brummell and Maclachlan 1989). An evolutionary shift of activity between closely related substrates within the GH45 might thus be due to a few amino acid substitutions, allowing to accommodate the decorations of xyloglucan into the substrate binding pocket of the enzyme.

G. viridula feeds on *Rumex spp*. which harbor constantly growing leaves where the proportion of primary cell wall content is likely to be very high. Taking into account that xyloglucan is the most abundant hemicellulose in the primary cell wall, having the ability to break down this polysaccharide would be highly beneficial for *G. viridula*. Thus, the biological need to acquire a xyloglucanase by altering the function of an existing enzyme is conceivable in an evolutionary perspective. Moreover, due to the restricted knowledge of PCWDEs in phytophagous beetles (and insects in general), future research will unravel whether neo- and bi-functionalization events occur frequently within glycoside hydrolase families. According to the CAZy database (Lombard et al. 2014) (http://www.cazy.org/), innovations in the enzymatic activity of glycoside hydrolase families seem to be the rule rather than the exception, e.g. GH1, GH3 or GH12. It is likely, however, that neo- and bi-functionalization events may have been required, as an evolutionary response to altering the composition of plant cell walls, ultimately promoting host plant adaptation.

According to previous work on insect-derived GH9s (Shelomi et al. 2016; Watanabe et al. 1998; Willis et al. 2011), it was surprising to find that the GH9 encoded by *G. viridula* was not active on any of the substrates we tested. Hence, this protein is either

active on a substrate we have not tested yet or it has lost its enzymatic activity. Although the former hypothesis seems more likely, our qPCR results demonstrated low to no expression in either gut tissue or the rest of the body of *G. viridula*, suggesting the possibility that the corresponding gene is currently being purged from the genome. In comparison, the expression of the genes encoding both GH45s in this species was high in the gut tissue and, together with our functional characterization, clearly indicated the digestive function of the corresponding proteins. Before the present study, there have been only two functionally characterized GH9 proteins in beetles: one was identified in the tenebrionid *T. castaneum,* which was characterized as an endo-β-1,4-glucanase (Willis et al. 2011), and the other in the cerambycid *A. glabripennis,* which did not exhibit any enzymatic activity under the conditions tested (McKenna et al. 2016). Thus, the two GH9 proteins analyzed in beetles of the superfamily Chrysomeloidea (see *A. glabripennis* (McKenna et al. 2016) and the present study) were both found to be inactive. In both cases, the loss of cellulolytic activity might have evolved in favor of the more efficient cellulolytically active enzymes of other families, i.e. GH45s. Until more GH9 proteins are discovered and characterized from other Chrysomeloidea species, this hypothesis remains highly speculative. Based on the data obtained (related to the *G. viridula* GH9 protein), the question arises whether what we observed here is a gene in the process of genomic “extinction”.

For insight into the role of the *G. viridula* GH45 proteins on a more natural substrate, we assessed whether the heterologously expressed GH45 proteins are active on the extracted plant cell wall of *R. obtusifolius.* The action of GH45-1 on this substrate revealed distinct breakdown patterns which are comparable in size with the corresponding breakdown products released by the action of GH45-1 on purified xyloglucan. This similarity suggests that xyloglucan is present in *R. obtusifolius,* and that GH45-1 has both access to and the ability to break it down directly from the cell wall network. In contrast and much to our surprise, incubating GH45-2 with extracted plant cell wall did not end up producing any breakdown products. Several hypotheses may explain these unexpected results: (i) cellulose is rarely or not at all present in *R. obtusifolius*, (ii) only crystalline cellulose is present within the extracted cell wall, (iii)

cellulose is not accessible by the enzyme due to its being packed compactly within other plant cell wall polysaccharides i.e. xyloglucan and pectins. The first hypothesis seems unlikely, as cellulose has been shown to be a constant constituent of the plant primary cell wall in all plants (Keegstra 2010). Regarding the second statement, our functional characterization clearly demonstrated that *G. viridula* GH45-2 is an endo-β-1,4-glucanase. It is widely accepted that endo-active cellulases are able to act only on the amorphous regions of cellulose but are useless on crystalline cellulose (as shown in Fig. 3B). Although it has been proposed that cellulose in the plant's primary cell wall has less crystallinity compared to the secondary wall (which would explain the presence of an endo-acting cellulase in *G. viridula* in the first place), the packing of cellulose within the primary cell wall remains elusive (Cosgrove 2014; Thomas et al. 2013). A recent article focusing on the packing of cellulose within cell walls of *A. thaliana* demonstrated that the cellulose in primary cell walls was in an almost entirely crystalline state, and as the secondary walls developed, there was an increase in the proportion of amorphous regions within cellulose microfibrils (Ruel et al. 2012). As the leaves of *R. obtusifolius* constantly grow, the proportion of primary cell walls compared to secondary walls should be high. Why, therefore, does *G. viridula* express a cellulase which mostly acts on amorphous regions? Hence, the third hypothesis seems to be the most likely. The access of cellulose within the primary cell wall is somewhat limited to enzymes and may require the synergistic effort of several PCWDEs rather than being accomplished by a single GH45 cellulase (Kostylev and Wilson 2012; Mansfield et al. 1999). Although co-incubation experiments with both *G. viridula* GH45s in our lab did not support this hypothesis, it is likely that other GH families encoded by *G. viridula* may contribute to the proposed synergism, e.g. GH5 and GH28s (Busch et al. 2017; Kirsch et al. 2014).

To validate the physiological importance of GH45 enzymes in *G. viridula*, we performed expression knockdown experiments of both GH45-encoding genes. We hypothesized that plant-cell-wall polysaccharides provide extensive energy storage which, if properly exploited, can be of important nutritive value for the insect. Fig. 1B clearly shows the ability of *G. viridula* gut content to degrade CMC to glucose, a process that is likely facilitated by the synergy of GH45-2 and β-glucosidases, both of

which are known to be ubiquitous in insects (Watanabe and Tokuda 2010). Based on the above hypothesis, knocking down the expression of the GH45-encoding genes should reduce or even prevent weight gain and even affect mortality in *G. viridula*. We were able to successfully down-regulate the expression of both GH45-encoding genes. Although the expression of the gene encoding GH45-2 was highly down-regulated in RNAi-treated insects (89.3 % on day four post-injection), the corresponding reduction of enzymatic activity reached only 26.2 %, suggesting that more than one cellulase is present in the larval gut of *G. viridula*. A CMC-based zymogram of *G. viridula* gut contents revealed the presence of at least five cellulase isozymes, indicating the presence of at least four other uncharacterized endo-cellulases in the larval gut. Unsurprisingly, we did not see any differences in weight gain or mortality in GH45-2 silenced insects compared to in a GFP control, as the other potential cellulases likely compensate for the reduction of the amount of GH45-2 protein due to the silencing. A similar lack of striking changes in phenotype was observed in GH45-1 silenced insects compared to controls, suggesting that the breakdown products of the individual cell wall polysaccharides may play a less important role in nutrient acquisition than anticipated. A lack of striking changes in the phenotype of RNAi-treated insects may have occurred because both GH45 enzymes may function as secondary enzymes by allowing the insect to access the plant cell content rather than being responsible for the exploitation of the plant cell wall polysaccharides for metabolic purposes, as has been suggested in our previous work (Busch et al. 2017). Consequently, the insect would not need to break down the recalcitrant polysaccharide network down to sugar monomers, but would gain access to many simple sugars and proteins from the plant cells. If this logic is true, and taking into account that protein-bound nitrogen is the limiting factor required for growth in phytophagous insects (Kainulainen et al. 1996; Kerslake et al. 1998; Rossi et al. 1996), reduced access to plant cells should result in striking phenotype differences compared to control animals. However, and as stated elsewhere (Busch et al. 2017), reduced xyloglucanase activity in GH45-1-silenced insects may be compensated for by the action of other glycoside hydrolase families present in *G. viridula,* such as GH5 mannanases (Busch et al. 2017) or GH28 pectinases (Kirsch et al. 2014). It is also

possible that the gut microbiota may contribute to the degradation of the plant cell wall in phytophagous beetles and may compensate for the reduction of the xyloglucanase and cellulase activity due to the silencing of the GH45 genes. However, examples of microbial symbionts within phytophagous beetles contributing to plant cell wall degradation are rare, and only one chrysomelid beetle is known to harbor a pectinolytic symbiont (Salem et al. 2017). Nonetheless, a lack of change of phenotype for GH45-1-silenced animals remains difficult to understand as xyloglucanases from plant pathogenic fungi were shown to be essential to insects' ability to penetrate and infest plant cells (Ma et al. 2017). Alternatively, a lack of a clear phenotype may have been caused by the optimal rearing conditions used for our experiments such as optimal food quality and supply. *Gastrophysa viridula* is an oligophagous which can feed, besides *R. obtusifolius* as main host plant, on other Polygonacious plants. However, it seems that full beetle development can only be achieved when feeding on *R. obtusifolius*. Choosing a non-optimal host plant for our experiments may have resulted in a more striking phenotype in silenced larvae.

In summary, the ability of a phytophagous beetle to efficiently degrade plant matter depends greatly on the set of PCWDEs the beetle's genome can express. The more complex this set is, the greater the synergism among PCWDEs and the more efficient the breakdown of the plant cell wall. As a consequence, the animal may take advantage of the released sugar monomers and, more important, improve its access to the nitrogen-rich plant cell content. Based on the present work and our previous studies, focusing on GH28 pectinases and GH5 mannanases (Busch et al. 2017; Kirsch et al. 2014), *G. viridula* possesses the ability to break down almost every component of the plant primary cell wall, i.e. pectins, mannans, amorphous cellulose and xyloglucan. Notably, the most abundant polysaccharides (cellulose and xyloglucan) are degraded solely by GH45 proteins, and these have likely evolved from a gene duplication event. We strongly believe that PCWDEs allowed beetles of the Phytophaga (Marvaldi et al. 2009) to thrive on plant tissue and may have been a prerequisite for the ability of these animals to adapt to their food.

Experimental procedures

Leaf beetles

Gastrophysa viridula larvae and adults were collected on *Rumex obtusifolius* in the vicinity of Jena (Germany) and brought to the lab. Individuals were reared in plastic containers on fresh *R. obtusifolius* leaves under a light/dark cycle of 16:8 hours at 18 °C. Insects were allowed to mate and oviposit, and the progeny was used for experimentation.

Preparation of regenerated amorphous cellulose (RAC)

RAC was prepared according to (Zhang et al. 2006). Briefly, 600 µl of double distilled water was added to 200 mg of microcrystalline cellulose (Avicel). Then, 10 ml of ice-cold ortho-phosphoric acid (86 %) was added while the mixture was stirred vigorously; after that, the solution was left for 60 min on ice and stirred occasionally. Ice-cold double-distilled water (10 ml) was applied four times with vigorous stirring in between. After centrifugation (10,000 g at 4 °C for 20 min), the pellet was washed three times by re-suspension in 40 ml ice-cold double-distilled water. Then, 40 ml of ice-cold 25 mM Na_2CO_3 was added, followed by three consecutive washing steps as described above until pH 5.0 - 7.0 was reached.

Protein purification and gel electrophoresis

After each purification step, fractions were deposited on a 1 % agarose plate containing 0.1 % carboxymethyl cellulose (CMC) in 40 mM citrate/phosphate buffer pH 5.0. CMCase activity was tracked after staining the plates with Congo red followed by destaining using 1 M NaCl, and cellulolytically active fractions were used for further purification steps.

One hundred *G. viridula* third-instar larvae were dissected in 40 mM citrate/phosphate buffer pH 5.0 containing a protease inhibitor cocktail (cOmplete, EDTA-free Protease Inhibitor Cocktail, Sigma-Aldrich). The guts were removed and pooled in 200 µl of the same buffer/inhibitor mixture containing 1% Triton X-100 and 50 mM NaCl. Guts were homogenized on ice using a 15 ml Dounce Tissue grinder (Wheaton, Millville, NJ, USA). The homogenate was centrifuged (100,000 g, 60 min, 4 °C) and the pellet discarded. Four times the volume of ice-cold acetone was added to the supernatant

and the mixture was then incubated for 10 min, followed by centrifugation (16,000 g, 4 °C and 10 min). The resulting protein pellet was dissolved in 1 ml Tris-HCl pH 8.0 containing 50 mM NaCl and a protease inhibitor cocktail (cOmplete, EDTA-free Protease Inhibitor Cocktail, Sigma-Aldrich). The protein mixture was then loaded on a 1 ml RESOURCE Q anion exchange chromatography column (GE Healthcare) attached to an Äkta FPLC System (GE Healthcare). After the column was washed thoroughly, bound proteins were eluted over 20 column volumes (1 ml) using a linear NaCl gradient ranging from 0 to 600 mM. Two fractions positive for CMCase activity were put in binding buffer (20 mM Tris-HCL, 0.5 M NaCl, 1 mM $MnCl_2$, 1mM $CaCl_2$, pH 7.4) using Zeba Desalt Spin Columns (Thermo Scientific) following the manufacturer's protocol. The resulting 2 ml sample was loaded on a HiTrap ConA 4B affinity chromatography column (GE Healthcare). After the column was washed thoroughly, bound proteins were eluted in six 1 ml fractions using elution buffer (0.5 M methyl-α-D-glucopyranoside, 20 mM Tris-HCl, 0.5 M NaCl, pH 7.4). Fractions positive for CMCase activity were adjusted to 20 mM Tris-HCl, pH 8.0 containing 0.15 M NaCl using Zeba Desalt spin columns. The resulting 5 ml sample was concentrated to 600 µl using Pierce concentrators with a 9 kDa molecular weight cutoff (Thermo Scientific). The concentrated sample was loaded on a Superdex 200 10/300 GL size-exclusion column (GE Healthcare), and proteins were eluted into 500 µl fractions using 20 mM Tris-HCl, pH 8.0 containing 0.15 M NaCl. Three hundred microliters of each fraction was precipitated using 1 % trichloroacetic acid and 0.02 % sodium deoxycholate. The resulting pellets were washed twice in 100 % ice-cold acetone, then dissolved and boiled in XT-sample buffer containing XT-reducing agent (Bio-Rad). Samples were then loaded on a Criterion XT gradient 4-12 % precast gel (Bio-Rad) and run for 2 h in XT-MOPS buffer at 120 V constant voltage. The gel was then fixed in 40% (v/v) ethanol, 7 % (v/v) acetic acid for 2 h followed by staining using colloidal Coomassie (Neuhoff et al. 1985).

In-gel digestion of proteins

Protein bands of interest were excised from the Coomassie-stained gels, cut into small pieces, washed several times with 25 mM NH_4HCO_3 and destained with 50 % acetonitrile (ACN) in 25 mM NH_4HCO_3. The proteins were then reduced with 10 mM

dithiothreitol (DTT) at 50 °C for 1 h and alkylated with 55 mM iodoacetamide (IAA) at room temperature in the dark for 45 min. Next, destained and washed dehydrated gel pieces were rehydrated for 60 min in a 12 ng/µL solution of porcine trypsin (Promega) in 25 mM NH_4HCO_3 at 4 °C and incubated overnight at 37 °C. The tryptic peptides were extracted from gel pieces in 75 % ACN/ 5 % formic acid (FA), and dried down in a SpeedVac (Shevchenko et al. 2006).

LC-MS/MS analysis

Each sample was injected onto a nanoAcquity nanoUPLC system (Waters) coupled to a Q-ToF HDMS mass spectrometer (Waters, Manchester, UK). Peptides were initially transferred with 0.1 % aqueous formic acid for desalting onto a Symmetry C18 trap-column (20 x 0.18 mm, 5 µm particle size) at a flow rate of 15 µL min^{-1} (0.1 % aqueous FA), and subsequently eluted onto a nanoAcquity C18 analytical column (200 mm x 75 µm ID, BEH 130 material, 1.7 µm particle size) at a flow rate of 350 nL/min with the following gradient: 1 to 30 % A (composed of 0.1 % FA and 100 % ACN) over 13 min, 30 to 50 % A over 5 min, 50 to 95 % A over 5 min, isocratic at 95 % A for 4 min, and returned to 1 % A over 1 min. The analytical column was re-equilibrated for 9 min prior to the next injection. The eluted peptides were transferred to the nanoelectrospray source of a Synapt HDMS tandem mass spectrometer (Waters, Manchester, UK) that was operated in V-mode with a resolution power of at least 10,000 full width at half maximum. All analyses were performed in positive ESI mode. A 650 fmol/µL human glu-fibrinopeptide B in 0.1 % FA/ACN (1:1 v/v) was infused at a flow rate of 0.5 µL per min through the reference Nano-LockSpray source every 30 s to compensate for mass shifts in MS and MS/MS fragmentation mode. LC-MS data were collected using data-dependent acquisition (DDA). The acquisition cycle consisted of a survey scan covering the range of m/z 400-1500 Da followed by MS/MS fragmentation of the four most intense precursor ions collected at 1 s intervals in the range of 50-1700 m/z. Dynamic exclusion was applied to minimize multiple fragmentations for the same precursor ions.

Data Processing and protein identification

DDA raw files were collected using MassLynx v4.1 software and processed using ProteinLynx Global Server Browser (PLGS) v2.3 software (Waters, Manchester, UK) under baseline subtraction, smoothing, de-isotoping, and lockmass-correction. The peptide fragment spectra were searched against a sub-database containing common contaminants (human keratins and trypsin). The following searching parameters were applied: fixed precursor ion mass tolerance of 15 ppm for survey peptide, fragment ion mass tolerance of 0.03 Da, estimated calibration error of 0.003 Da, one missed cleavage, fixed carbamidomethylation of cysteines and possible oxidation of methionine. Spectra that remained unmatched by database searching were interpreted *de novo* to yield peptide sequences. A 0.002 Da mass deviation for *de novo* sequencing was allowed, and sequences with a ladder score (percentage of expected y- and b-ions) exceeding 40 were subjected for homology-based searching using the MS-BLAST program (Shevchenko et al. 2001) installed on an in-house server. MS-BLAST searches were performed against the NCBInr database, sub-database insect (downloaded on October 21, 2014) and the *in silico* translated *G. viridula* gut transcriptome using the following settings: scoring table, 100; filter, none; expect, 100; matrix, PAM30MS; advanced options, no-gap-hspmax100-sort_by_totalscore-span1. Statistical significance of the matched hits was evaluated according to the MS-BLAST scoring scheme (Shevchenko et al. 2001). In parallel, pkl files of MS/MS spectra were generated and searched against NCBInr database (updated on October 20, 2014) combined with the *G. viridula* sub database using MASCOT software version 2.4 (searching parameters are described above). Hits were considered to be confident if at least three peptides were matched with ion scores above 30, or proteins were identified by one or two peptides with score of 55 or better.

Full-length gene amplification using degenerate primers

To amplify the cDNA corresponding to the GH45 (GH45-2) identified after protein purification, degenerate primers were designed (Table 1) based on recently described GH45 sequences of various leaf beetle species (Pauchet et al. 2010), excluding *G. viridula* GH45-1. cDNAs initially generated for the rapid amplification of cDNA ends

PCR (RACE-PCR) as described by (Pauchet et al. 2010) were used as the template, and the PCR reactions were performed using the Avantage 2 PCR kit (Clontech). Amplicons of estimated appropriate size were gel-purified and cloned into PCR4-TOPO (Invitrogen). TOP10 chemically competent *Escherichia coli* cells (Invitrogen) were transformed and incubated overnight on a LB agar plate containing ampicillin (100 µg/ml). Clones were randomly picked and cultured in DYT medium containing 100 µg/ml ampicillin. Plasmid isolation was performed in 96 deep-well plate using the 1-step DNA Isolation Kit for plasmids (Nexttec) on a Freedom EVO platform (Tecan). The recombinant DNA was sequenced in both directions using Sanger sequencing, and the resulting sequences were blasted against the NCBInr database. Amplicons with confirmed GH45 hits were used as a template to design gene-specific primers (Table S1), and full-length cDNA sequences were amplified by 5´- and 3´-RACE-PCRs using RACE-ready cDNAs as described by (Pauchet et al. 2010). Fragments obtained after RACEs-PCR were cloned into pCR4-TOPO and Sanger sequenced as described above.

Transient protein expression in insect cells

Open reading frames (ORFs) of target genes were amplified from cDNAs using gene-specific primers designed according to either *G. viridula* transcriptomic data (GH45-1 and GH9) or based on RACE-PCR results as described above (GH45-2). The forward primer was designed to introduce a Kozak sequence at the beginning of the ORF, and the reverse primer was designed to omit the stop codon. RACE-ready cDNAs as described above were used for amplification, and PCR reactions were performed using a high-fidelity Taq polymerase (AccuPrime, Invitrogen). PCR products were cloned into pIB/V5-His TOPO (Invitrogen) in frame with a V5-$(His)_6$ epitope. Top10 chemically competent *E. coli* cells were transformed and treated as described above. To select constructs for which the sequence of interest had ligated in the correct direction, randomly picked colonies were checked by colony PCR using the OpIE2 forward primer (located on the vector) and the gene-specific reverse primer (Table S1). Positive clones were further cultured in DYT medium containing 100 µg/ml ampicillin. After plasmid isolation using GeneJET Plasmid Miniprep Kit (Thermo Scientific), the recombinant plasmids were sequenced (see above) to confirm whether

the ORF had been correctly inserted into the vector. Positive constructs were then transfected in *Sf*9 cells (Invitrogen) using FuGENE HD (Promega) as a transfection reagent. First, successful expression was determined by transiently transfecting three clones per construct in a 24-well plate format. After 72 h, the culture medium was harvested, and successful expression was verified by Western blot using the anti-V5-HRP antibody (Invitrogen). In order to collect enough material for downstream enzymatic activity assays, a single clone per construct was chosen to be transiently transfected in a 6-well plate format. After 72 h, culture medium was harvested and centrifuged (16,000 x g, 5 min, 4 °C) to remove cell debris; finally the medium was stored at 4 °C until further use. Again, successful expression was verified by Western blot using the anti-V5-HRP antibody. The enzymatic activity of recombinant proteins was initially tested by agarose diffusion assays using CMC as a substrate as described above.

Analysis of hydrolysis reaction products by thin layer chromatography (TLC)

The culture medium of transiently transfected cells was dialyzed and desalted as described in (Busch et al. 2017). Gut content samples were dialyzed three times against 500 ml double-distilled water within 24 h. The following substrates were tested: CMC, RAC, Avicel, filter paper, glucomannan, galactomannan, xyloglucan, glucomannan, galactomannan, polygalacturonic acid (all from Megazyme), xylan, citrus pectin, apple pectin, esterified pectin (all from Sigma Aldrich), beet pectin (kindly provided by CP Kelco, Grossenbrode, Germany) and plant cell wall extracted from *R. obtusifolius* as previously reported (Busch et al. 2017). Additionally, we tested the cello-oligomers D-(+)-biose to D-(+)-hexaose, the manno-oligomers D-(+)-biose to D-(+)-hexaose, dimer and trimer of xylan (all from Megazyme), and the di- and tri-galacturonic acid (Santa Cruz Biotech, Dallas, TX, USA) at a final concentration of 250 ng/µl per reaction. Samples were incubated and analyzed as previously described (Busch et al. 2017). The reference standard contained 2 µg of each oligomer: glucose, cellobiose cellotriose, cellotetraose and cellopentaose and mannose, mannobiose, mannotriose, mannotetraose and mannopentaose (all from Megazyme) and xylose (Sigma-Aldrich), xylobiose and xylotriose (both from Megazyme) and mono-, di- tri- galacturonic acid (all from Sigma-Aldrich).

pH optima and temperature optima

To test the temperature optima, dialyzed and desalted crude enzyme extracts were incubated with 0.5 % (w/v) CMC (GH45-2), or xyloglucan (GH45-1) in 20 mM citrate phosphate buffer (pH 5.0) at temperatures ranging from 20 °C to 80 °C in steps of 10 °C. In detail, each enzyme assay was performed with 24 µl dialyzed and desalted enzyme extract, 30 µl of 1% (w/v) substrate solution and 6 µl of 20 mM citrate phosphate buffer pH 5.0. Negative controls were carried out with 24 µl of distilled water instead of enzyme. The enzymatic activity was assayed for 45 min (GH45-1) and 120 min (GH45-2). The amount of reducing sugars produced in these reactions was measured using the dinitrosalicylic acid (DNS) method according to (Kirsch et al. 2014). To test for pH optima, dialyzed and desalted enzyme extracts were incubated at 40 °C (GH45-1) and 50 °C (GH45-2) with their respective substrates, as described above, and assayed in 20 mM citrate phosphate buffer ranging from pH 2.0 to 8.0 as well as in 50 mM sodium carbonate buffer covering pH 9.0 and 10.0. The amount of reducing sugars produced in these reactions was measured using the DNS method as described above. To test for thermal stability, dialyzed and desalted enzyme extracts were incubated without substrate for 16 h at different temperatures starting with the enzymes' optimum up to 80 °C in steps of 5 °C. Then, each enzyme was incubated with its respective substrate at pH 5.0 for 45 min and analyzed as described above. Each reaction was carried out in triplicate.

Preparation of double-stranded RNA and off-target prediction

Primers for the generation of double-stranded RNA (dsRNA) were designed for *G. viridula* GH45-1 and GH45-2, and for GFP as the control, yielding two 200 bp fragments and a 379 bp fragment, respectively. Off-target prediction and dsRNA synthesis were performed as described previously (Busch et al. 2017).

Injection of dsRNA and assessment of RNAi efficiency

Early second-instar *G. viridula* larvae were injected dorsally with 50 nl (150 ng) of target dsRNA into the metathorax, using a Nanoliter 2010 Injector (World Precision Instruments) attached to a three-dimensional micromanipulator, and were then put onto fresh *R. obtusifolius* leaves. To record weight gain and mortality, five animals per

replicate were injected with a total of six replicates for each target gene. To analyze gene expression and enzymatic activity, three animals per replicate were injected with a total of six replicates for each target gene. In addition to larvae injected with dsRNA targeting GFP, a non-injected control was also included. For quantitative PCR and enzymatic activity analyses, larvae were collected on days 1, 4 and 8 post-injection. Whole larvae were crushed in liquid nitrogen; half of the resulting powder was used for total RNA preparation and the other half was used for protein extraction and enzymatic assays.

Total RNA was isolated using the innuPREP RNA Mini Kit (Analytik Jena), following the manufacturer's protocol. The resulting RNA was then subjected to DNase digestion (Ambion), and its quality was subsequently checked using the RNA 6000 Nano LabChip kit on a 2100 Bioanalyser (both Agilent Technologies). Total RNA was used as a template to synthesize cDNAs using the Verso cDNA synthesis kit (Thermo Scientific). The resulting cDNA samples were then used for real-time quantitative PCR (qPCR) experiments, which were performed in 96-well hard-shell PCR plates on the CFX Connect Real-Time System (both Bio-Rad). All reactions were carried out using the 2-Step QPCR SYBR Kit (Thermo Scientific), following the manufacturer's instructions. Primers were designed using Primer3 (version 0.4.0) (Table S1). The specific amplification of each transcript was verified by dissociation curve analysis. A standard curve for each primer pair was determined in the CFX Manager (version 3.1) based on Cq-values (quantitation cycle) of qPCRs run with a dilution series of cDNA pools. The efficiency and amplification factors of each qPCR, based on the slope of the standard curve, were calculated using an integrated efficiency calculator of the CFX manager software (version 3.1). Ribosomal protein S3 (RPS3), extracted from our *G. viridula* larval gut transcriptome (Pauchet et al. 2010), was used as a reference gene, and the abundance of GH45 transcripts was expressed as RNA molecules per 1000 RNA molecules of RPS3.

To directly compare GH45-1/GH45-2 transcript abundance to GH45-1/GH45-2 enzymatic activity in RNAi-treated *G. viridula*, powdered material was suspended in 40 mM citrate/phosphate buffer at pH 5.0 containing a protease inhibitor cocktail

(cOmplete, EDTA-free Protease Inhibitor Cocktail, Sigma-Aldrich). Then, the samples were centrifuged (10 min, 16000xg, 4 °C), and the supernatant was collected and stored at 4 °C until further use. Protein concentration was estimated by Bradford protein assay (Bio-Rad). Enzymatic activity assays were carried out using the DNS method as described above, using 5 µg of extracted proteins in the reaction. Alternatively, 2.5 µg (GH45-2) and 5 µg (GH45-1) of total extracted proteins were prepared for zymogram analysis by diluting the sample in Laemmli buffer without any reducing agent. Samples were run on a 12.5% SDS-PAGE gel containing 0.1 % (w/v) of either CMC or Xyloglucan. Electrophoresis was carried out at 4 °C using pre-chilled running buffer. Gels were then washed three times in a 2.5 % Triton X-100 solution for 15 min each at 4 °C, before being equilibrated in the reaction buffer (50 mM citrate/phosphate buffer pH 5.0) for 16 h at 4 °C, followed by a 1 h incubation at 40 °C. The gels were then incubated in a 0.1 % (w/v) Congo red solution before being destained in 1 M NaCl until pale activity zones appeared against a dark red background.

Two life history traits were recorded after larvae were injected with dsRNA. First, larvae (we used groups of five insects per replicate, six replicates in total) were weighed on day 1 and day 8 post-injection. Then, growth rate was calculated using the formula “growth rate= log10(final weight)-log10(initial weight)/time (days)”. Finally, mortality was recorded at the end of the experiment.

Tissue-specific gene expression

Late-instar *G. viridula* larvae, actively feeding on leaves of *R. obtusifolius*, were used for total RNA extraction. Larvae were cut open from abdomen to head, and the complete gut was removed and stored separately from the rest of the body. Dissection and storage were carried out in RL solution (Analytik Jena). Three biological replicates were sampled, each containing three larvae. RNA extraction, generation of cDNAs and subsequent real-time qPCR experiments were performed as described above. Primers used for these experiments are listed in Table S1.

Statistical analysis

If not otherwise stated, data were analyzed in R version 3.2.0 (R-Development-Core-Team 2015). Statistical analyses of gene expression over time were performed as follows: The influence of GH45-1 and GH45-2 RNAi treatments (i45-1 and i45-2) over time (RNAi treatment and time used as categorical explanatory variables) on GH45-1 and GH45-2 transcript abundance was investigated using the generalized least squares method (gls from the nlme library (Pinheiro et al. 2015)) to account for the variance heterogeneity among the residuals. The varIdent variance structure was used, with a different variance for the combination of treatment and time (varIdent (form = ~1|combination of [treatment and time])). The influence of the explanatory variables was determined by sequentially removing explanatory variables starting with the full model and comparing the simpler model to the more complex one, using a likelihood ratio test (Zuur et al. 2009). Differences between factor levels were determined by factor level reduction (Crawley 2013). The influence of RNAi treatment on the enzymatic activity over time was analyzed either by the gls method as described above (GH45-1) or using a two-way ANOVA (GH45-2). Statistical details can be found in supplementary table S4. The Tukey HSD test was performed in order to find differences among the groups. To compare weight gain over time in RNAi-treated larvae, we calculated the relative growth rate for a period of 8 days and analyzed the data using a one-way ANOVA. Differences in mortality were analyzed using the test for the equality of proportions. Differences in tissue-specific gene expression were analyzed in SigmaPlot Version 11.0 using paired t-tests.

Acknowledgments

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch for technical support. We thank Emily Wheeler, Boston, for editorial assistance. We are also thankful to Roy Kirsch and David G. Heckel for their input on experimental design and for fruitful discussions. This work was supported by the Max-Planck-Gesellschaft.

Supplementary Material

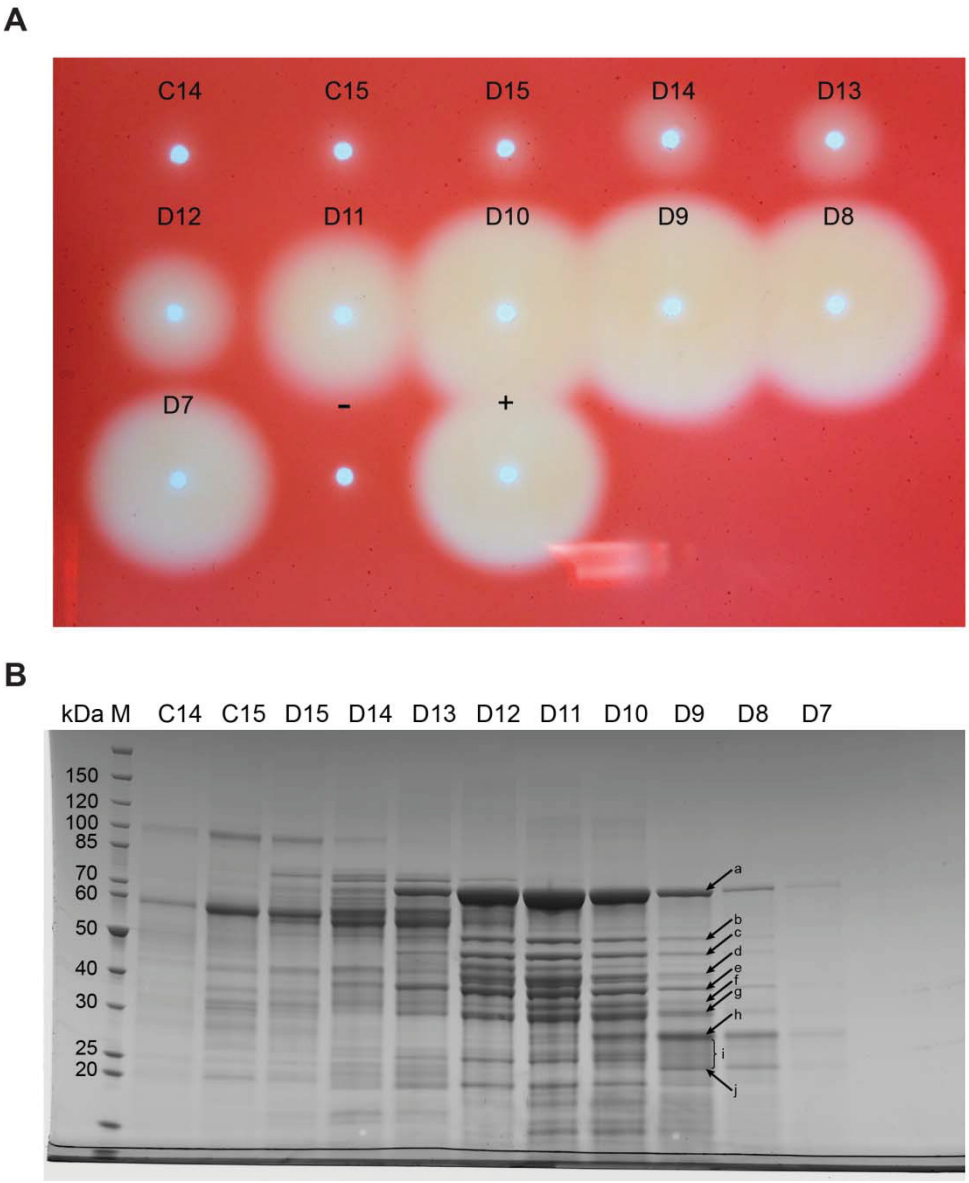


Fig. S1. Purification of cellulolytic proteins from larval guts. (A) After size-exclusion chromatography, all fractions were loaded on an agarose diffusion assay containing 0.1 % CMC. The assay was incubated for 16 h at 40 °C and pH 5.0 and subsequently stained using Congo red. (B) After TCA precipitation, each fraction was applied to an SDS-gel and stained using colloidal Coommassie. Fraction D9 was chosen for LC-MS/MS analysis, as it shows the highest cellulolytic activity. C14 to D7 indicates the fractions in which the purified samples were collected from the FPLC. Each arrow/letter (a-j) corresponds to samples analyzed by LC-MS/MS. + = positive control; - = negative control; kDa = kilodalton; M = marker

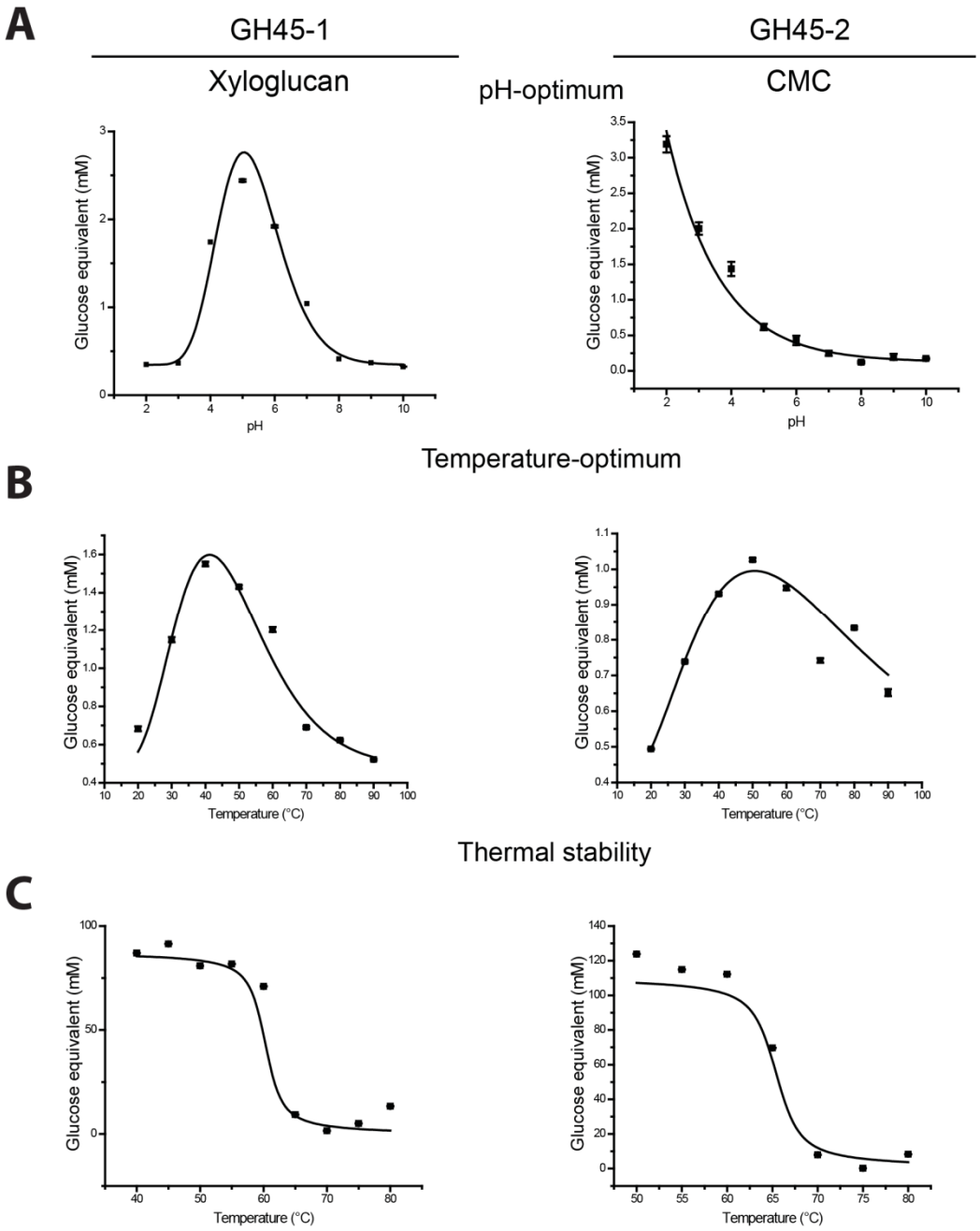


Fig. S2. figure legend see next page

Fig. S2. Analysis of optimal pH and temperature as well as thermal stability of GH45 proteins. Assays were performed using dialyzed and desalted cell culture medium containing target recombinant proteins. Enzymatic activity was determined by analyzing the amount of reducing sugars set free and was converted into millimolar (mM) of sugar monomer equivalents. (A) Target proteins were incubated with their respective substrate, and enzymatic activity was determined at several pH values ranging from 2.0 – 10.0. (B) Target proteins were incubated at temperatures ranging from 20 – 90 °C and the respective enzymatic activity recorded. (C) Target proteins were incubated without substrate for 16 h at temperatures starting with their respective optima and increasing to 80 °C. Then, the respective substrates were added and the proteins were incubated once more at their optimal temperature. Residual enzymatic activity was recorded and compared to a control. CMC = carboxymethyl cellulose; XylG = xyloglucan.

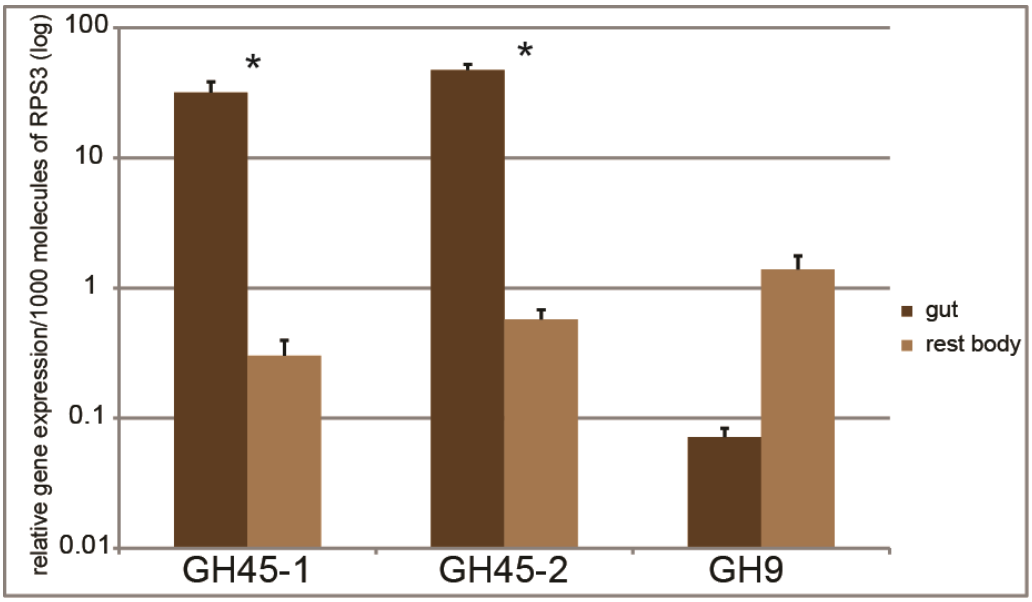


Fig. S3. Expression of beetle-derived GH9 and GH45 genes comparing gut tissue to the rest of the body. Third-instar actively feeding larvae were dissected in groups of three (three replicates per group). Gut tissue and rest of the body were used for total RNA extraction and subsequent quantitative RT-PCR. Gene expression was expressed as the copy number of target GH gene per 1000 molecules of RPS3 (control gene). Data were plotted using a log-transformed scale. Statistics were carried out using a paired t-test (statistical values see S1 table). Significant differences are indicated by an asterisk ($P \leq 0.05$).

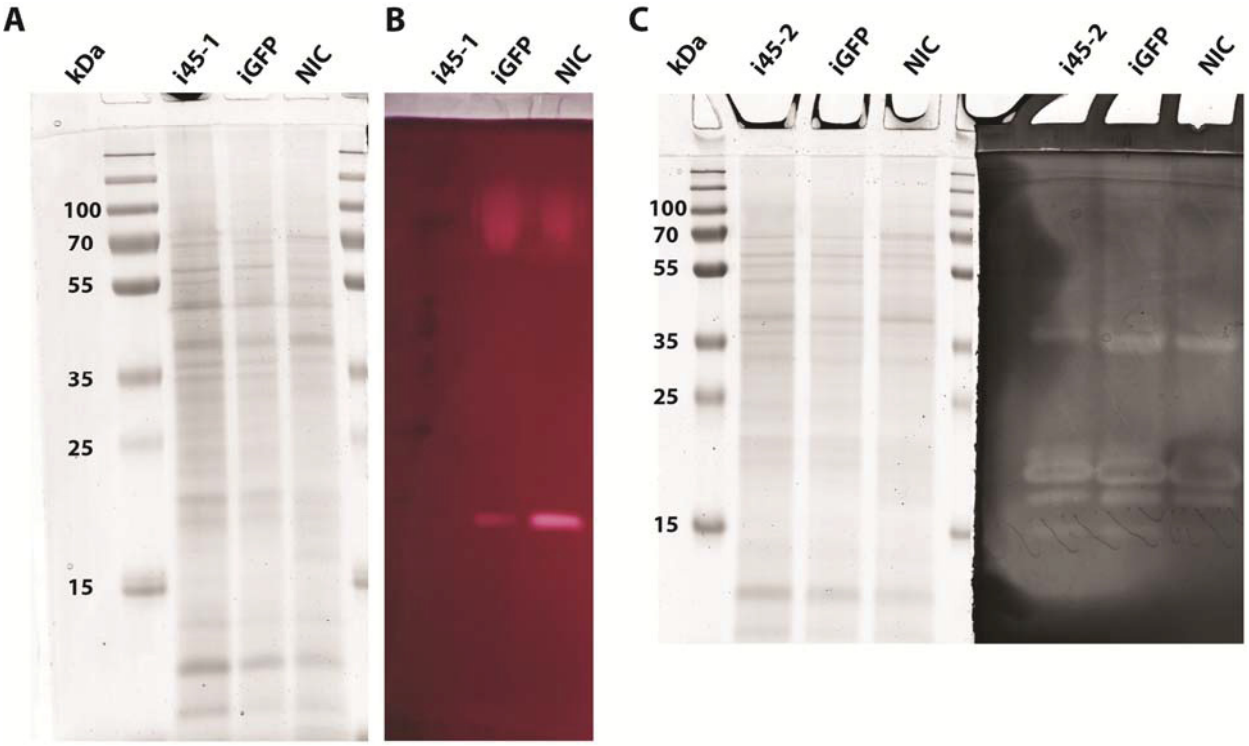


Fig. S4. Zymogram testing the activity of gut content against xyloglucan and CMC after RNAi treatment. Samples from larvae collected on day four post-injection as used in Figure 4 were applied to the zymograms (A) Protein loading control of larvae treated with dsRNA against GH45-1 (i45-1) and GFP (iGFP) including a non-injected control (NIC). (A) Five micrograms of total protein content was applied to a 0.1 % xyloglucan-containing semi-native SDS-PAGE gel and subsequently stained with Coomassie. (B) The other half of the same semi-native SDS-PAGE gel, after protein renaturation, was used to detect xyloglucanase activity. The gel was incubated at 40°C for one hour and activity bands subsequently revealed by Congo red staining. (C) Semi-native SDS-PAGE gel containing 0.1 % CMC. Total protein content of larvae treated with dsRNA against GH45-2 (i45-2) and GFP as well as a non-injected control was loaded (5 µg of total protein content) with subsequent Coomassie staining. (D) 2.5 µg total protein content was loaded and proteins were renaturated after the run. The gel was incubated at 40 °C for one hour and activity zones subsequently revealed by Congo red staining.

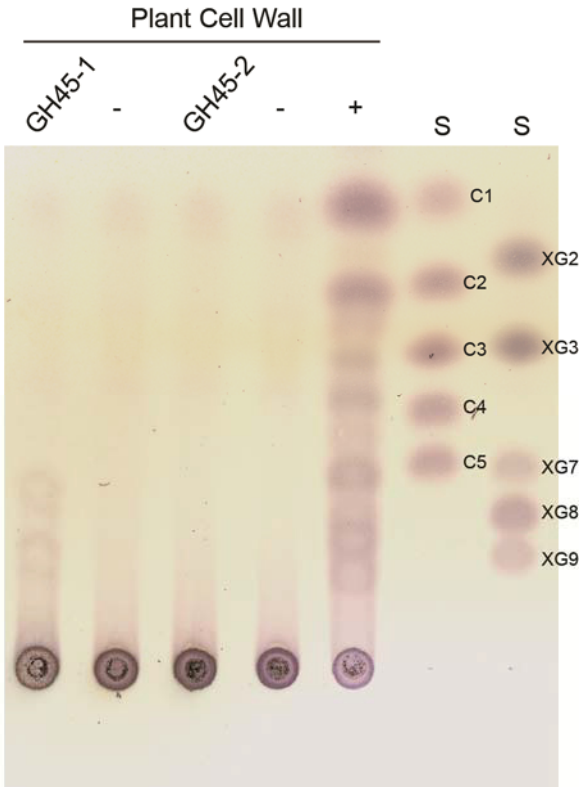


Fig. S5. Analysis of breakdown products of *Rumex obtusifolius* primary cell wall extracts incubated with *G. viridula* recombinant GH45 proteins. Thin layer chromatography of heterologously expressed GH45s when incubated with plant primary cell wall extract for 16 h at 40 °C and pH 5.0. Breakdown products were visualized by spraying 0.2 % orcinol (w/v) with methanol/sulphoric acid. Standards (S) used: C1-C5 = glucose to cellopentaose; XG2 - XG9 = isoprimeverose to xyloglucan nona saccharide. (-) represents negative controls where plant cell wall extracts were incubated with the culture medium of mock-transfected *Sf*9 cells.

3.3. Manuscripts III

Functional analyses of the horizontally acquired Phytophaga glycoside hydrolase family 45 (GH45) proteins reveal distinct functional characteristics

André Busch[1], Etienne G.J. Danchin[2] and Yannick Pauchet[1]

[1]Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany and [2]INRA, Université Côte d'Azur, CNRS, ISA, Sophia Antipolis, France.

Manuscript in preparation and to be submitted to

Molecular Biology and Evolution

Abstract

Cellulose, a major polysaccharide of the plant cell wall, consists of β-1,4-linked glucose moieties forming a molecular network recalcitrant to enzymatic breakdown. Although cellulose is potentially a rich source of energy, the ability to degrade it is rare in animals and was believed to be present only in cellulolytic microbes. Recently, it has become clear that some animals encode endogenous cellulases belonging to several glycoside hydrolase families (GHs), including GH45. GH45s are distributed patchily among the Metazoa and, in insects, are encoded only by the genomes of Phytophaga beetles. This study aims to understand both the enzymatic properties and the evolutionary history of GH45s in these beetles. To this end, we tested the enzymatic abilities of 37 GH45s derived from five species of Phytophaga beetles and learned that beetle-derived GH45s degrade three different substrates: amorphous cellulose, xyloglucan and glucomannan. Our phylogenetic and gene structure analyses indicate that at least one gene encoding a putative cellulolytic GH45 was present in the last common ancestor of the Phytophaga, and that GH45 xyloglucanases evolved several times independently in these beetles. The most closely related clade to Phytophaga GH45s contained fungal sequences, suggesting this GH family was acquired by horizontal gene transfer from fungi. Other than in insects, arthropod GH45s do not share a common origin and appear to have emerged at least three times independently.

Keywords: Chrysomeloidea; Curculionoidea, GH45, cellulase, xyloglucanase

Introduction

The major source of energy for most organisms on earth is D-glucose. Through photosynthesis, plants have evolved the ability to biosynthesize organic D-glucose from inorganic carbon dioxide. A surplus of electromagnetic energy has provided plants with a nearly unlimited access to glucose, giving them a reservoir for storing energy as well as access to material for building structural components during plant growth. The plant cell wall (PCW) consists of several glucose-derived polysaccharides, which form a protective wall against biotic and abiotic stresses. Traditionally, three kinds of polysaccharides are used as structural elements in the PCW: cellulose, hemicellulose and pectin. While the latter two are characterized by a variety of differently organized heteropolysaccharides, cellulose is a homopolymer and consists of β-1,4-linked anhydroglucose units forming a straight-chain polysaccharide. Through hydrogen bonding, individual chains attach to each other and form a resilient (para)crystalline structure (Chang 1981). On surface areas, cellulose is believed to organize itself into a state of low crystallinity, referred to as an “amorphous” state (Ruel et al. 2012). Depending on the developmental stage of the plant cell, the PCW is organized as follows: (i) the primary cell wall, which contains low amounts of crystalline cellulose (and surrounds plant cells in development) or (ii) the secondary cell wall, which comprises large amounts of crystalline cellulose (Cosgrove 2014). However, how native cellulose is organized in primary and secondary cell walls with regard to the ratio of amorphous to crystalline cellulose is still unclear (Knox 2008; Saxena 2007).

As the most abundant biopolymer on earth (Bayer et al. 1998), cellulose represents an abundant energy supply for any organism which has the ability to exploit it. Curiously, cellulose degradation has evolved only in few branches of the tree of life. Until the end of the 20th century, cellulose degradation was only known to be performed by microorganisms such as plant pathogenic bacteria (Chambost J.P. 1987; Py et al. 1991), saprophytic fungi (Schulein 1997) or mutualistic symbionts in insects and ruminants (Breznak and Brune 1994; Rincon et al. 2001). However, in 1998, the first endogenous cellulases of animal origin were identified in cyst

nematodes found in parasitic plants (Smant et al. 1998) and termites (Watanabe et al. 1998). Several other independent discoveries of cellulases in a variety of Metazoa followed, and to date endogenous cellulases encompass the phyla Arthropoda, Mollusca and Nematoda (Faddeeva-Vakhrusheva et al. 2016; Girard and Jouanin 1999; Kikuchi et al. 2004; Pauchet et al. 2010; Sakamoto and Toyohara 2009).

Cellulases are conventionally classified according to their mode of action. Endo-β-1,4,-glucanases (EC 3.2.1.4) break down cellulose by releasing randomly sized cellulose fragments and are known to act only on amorphous cellulose. Cellobiohydrolases (exo-β-1,4,-glucanases; EC 3.2.1.91) degrade cellulose from its terminal regions by releasing cellobiose and occasionally cellotriose. In microbes, cellobiohydrolases were shown to degrade amorphous as well as crystalline cellulose (Takahashi et al. 2010). Finally, cellobiosidases (β-glucosidases; EC 3.2.1.21) accept the released cellobiose as substrate and convert it into glucose. All three types of cellulases act synergistically and are necessary to degrade the cellulosic network efficiently (Kostylev and Wilson 2012).

Cellulases are distributed into 14 of the 156 currently described families of glycoside hydrolases (GHs), according to the carbohydrate-active enzyme (CAZy) database (www.cazy.org, (Lombard et al. 2014). Assignment to different GH families is based on sequence similarities. The best-described cellulolytic GH families encompass GH5 (Aspeborg et al. 2012) and GH9 (Watanabe and Tokuda 2010). Together with GH45s, GH5s and GH9s are found to be encoded by the genome of some insects (Keeling et al. 2013; McKenna et al. 2016; Pauchet et al. 2014b; Schoville et al. 2018; Vega et al. 2015). Based on our previous work, GH45s are commonly distributed in the Phytophaga clade of beetles (Marvaldi et al. 2009), which encompasses the superfamilies Chrysomeloidea (leaf beetles and longhorned beetles) and Curculionoidea (weevils and bark beetles) (Kirsch et al. 2012; Pauchet et al. 2014a; Pauchet et al. 2010). The first GH45 that was functionally characterized in a beetle originated from *Apriona germari* (Chrysomeloidea: Cerambycidae: Lamiinae), which had the ability to degrade amorphous cellulose (Lee et al. 2004). Until recently, GH45s in beetles have been functionally characterized in only a few Chrysomeloidea

species, mostly Cerambycidae (Chang et al. 2012; Mei et al. 2015; Pauchet et al. 2014a) and *Diabrotica virgifera virgifera* (Chrysomelidae: Galerucinae) (Valencia et al. 2013), and another two from our previous study in *Gastrophysa viridula* (Chrysomelidae: Chrysomelinae) (Busch et al. 2018a). Although GH45 sequences have been identified in Curculionoidea beetles (Keeling et al. 2013; Pauchet et al. 2010; Vega et al. 2015), to date none has ever been functionally characterized.

Interestingly, GH45s are not only found in multicellular organisms but are widely encoded by microbes (Davies et al. 1995; DeBoy et al. 2008; O'Connor et al. 2014; Sheppard et al. 1994). The distribution of this gene family in the Metazoa appears to be patchy and has so far been recorded only in a few species within the phyla Mollusca (Rahman et al. 2014; Sakamoto and Toyohara 2009; Xu et al. 2002), Nematoda (Kikuchi et al. 2004; Palomares-Rius et al. 2014); GH45s have also been recorded in Arthropoda (Song et al. 2017). If GH45s had evolved in the last common ancestor (LCA) of the Metazoa and subsequently been inherited by their descendants, we would expect the patchy distribution of GH45 genes observed within the Metazoa to be due to multiple independent losses. If this hypothesis were true, phylogenetic analyses would recover metazoan GH45s as a single monophyletic clade. However, two previous studies focusing on the evolutionary origin of GH45s in nematodes and mollusks have suggested instead that GH45s were acquired from a fungal donor by horizontal gene transfer (HGT) (Kikuchi et al. 2004; Sakamoto and Toyohara 2009). The first attempt to clarify the evolutionary history of GH45s in beetles also proposed an HGT from a fungal source but was unable to reach definite conclusions (Calderon-Cortes et al. 2010) because of the low number of sequences used for the phylogenetic analysis. A more comprehensive approach followed in 2014 (Eyun et al.), which included more GH45 sequences. Still, the variety of GH45 sequences in the latter study was poor, resulting in a similarly elusive outcome. Thus, the evolutionary history of GH45s appears to be complex, and their inheritance in beetles remains enigmatic.

Therefore, the major aim of our study was to trace the evolutionary origin of the GH45 family within the Phytophaga and to analyze how the function of the corresponding

proteins evolved in this large group of beetles. Based on previous research on the ancestral origin of GH45s (Kikuchi et al. 2004; Sakamoto and Toyohara 2009), we hypothesize that an HGT event occurred at one or more stages of the evolution of the Phytophaga. Additionally, we analyzed other Arthropods, including Oribatida and Collembola, as well as several non-arthropod species, including Nematoda, Tardigrada and Rotifera. In this study, we combined functional and phylogenetic analyses to unravel the origin and evolution of the GH45 family in Phytophaga beetles. We first functionally characterized 37 GH45s from five beetle species -- four beetles of the Chrysomelidae (leaf beetles) and a beetle of the Curculionidae (weevils) -- to determine whether these GH45s harbor cellulase activity, and whether they may have evolved other functions. We then combined these functional data with amino acid alignments of the GH45 catalytic sites to pinpoint amino acid substitutions which might lead to substrate shifts. Finally, we performed phylogenetic analyses to ask (i) how many GH45 genes were present in the LCA of the Phytophaga and (ii) whether this gene family is ancestral in the Metazoa or, instead, acquired by HGT. The aim of our study was to provide the first comprehensive overview regarding the evolution in beetles of the GH45 family and to assess the role of these genes in the evolution of herbivory.

Results

Functional analyses of the Phytophaga GH45 proteins reveal distinct enzymatic characteristics

Our previous transcriptome analyses (Eyun et al. 2014; Kirsch et al. 2012; Pauchet et al. 2010) revealed a set of endogenous GH45 genes distributed within several beetles of the superfamilies Chrysomeloidea and Curculionoidea. We investigated the product of GH45 genes from four beetle species belonging to the family Chrysomelidae, namely, *Chrysomela tremula* (CTR), *Phaedon cochleariae* (PCO), *Leptinotarsa decemlineata* (LDE) and *Diabrotica virgifera virigfera* (DVI), and from one species belonging to the family Curculionidae, the rice weevil *Sitophilus oryzae* (SOR), for a total of 33 GH45 sequences (Table S1). By re-examining the corresponding transcriptomes as well as the recent draft genome of *L. decemlineata* (Schoville et al. 2018), we identified four extra GH45 sequences (Table S1). The resulting 37 GH45s were successfully expressed in *Sf*9 insect cells (Fig. 1A). All GH45s had an apparent molecular weight of ~35 kDa (Fig. 1A). The increase in molecular size compared to the expected size (~25 kDa) was likely due to post-translational N-glycosylations as well as to the artificially added V5/$(His)_6$-tag.

To explore the cellulolytic capabilities of these proteins, we first applied crude *Sf*9 culture medium containing individual recombinant GH45s to agarose plates supplemented with 0.1 % carboxymethyl cellulose (CMC) (Fig. 1B). Activity halos were visible for at least two GH45s per target species. The intensity of the observed activity halos varied from large clearing zones (for example, PCO4 or LDE2) to small or medium ones (such as PCO3 or LDE7). These differences were likely due to the catalytic efficiency of each individual GH45 as well as to the concentration of the crude protein extracts we used. Each clearing zone, independent of its intensity and size, indicated endo-β-1,4-glucanase activity. To further assess enzymatic characteristics of these GH45s, we performed assays with a variety of plant cell wall-derived polysaccharides as substrates and analyzed the resulting breakdown products on thin layer chromatography (TLC) (Table 1; Fig. S1 to S5).

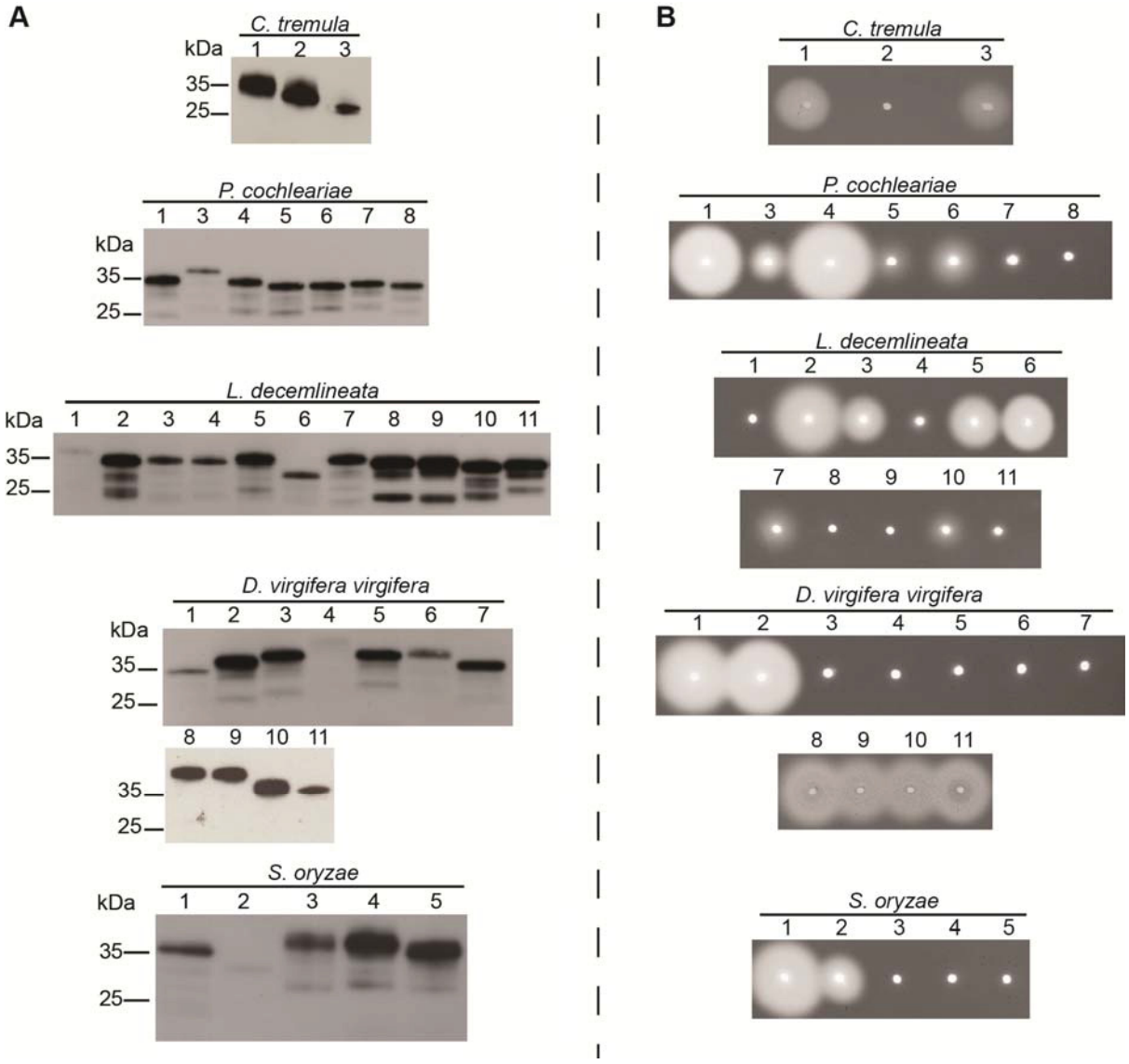


Fig. 1 Western Blot and CMC-based agarose-diffusion-assay of target GH45 proteins. A) Western Blot of target recombinant enzymes expressed in frame with a V5/His6 after heterologous expression in insect Sf9 cells. After 72 h crude culture medium of transfected cells were harvested and analysed by Western blotting using an anti-V5-HRP coupled antibody. B) Crude culture medium of transfected cells were applied to an agarose-diffusion-assay containing 0.1 % CMC. Activity halos were revealed after 16 h incubation at 40 °C using Congo red. Numbers above Western blot and agarose-diffusion-assay correspond to the respective species GH45s depicted in supplementary table S1.

Table 1: GH45 enzymatic activity observed on TLC.

	CMC	RAC	CC	C3	C4	C5	C6	Glu	Gal	XG
CTR1	+	+	-	-	-	-	+	+	-	-
CTR2	-	-	-	-	-	-	-	+	-	+
CTR3	+	+	-	-	+	+	+	+	-	-
PCO1	+	+	-	-	-	+	+	-	-	-
PCO3	+	+	-	-	-	-	+	+	-	-
PCO4	+	+	-	-	+	+	+	+	-	-
PCO5	-	-	-	-	-	-	+	+	-	-
PCO6	+	-	-	-	-	-	+	+	-	-
PCO7	-	-	-	-	-	-	-	-	-	+
PCO8	-	-	-	-	-	-	-	-	-	-
LDE1	-	-	-	-	-	-	-	-	-	-
LDE2	+	+	-	-	+	+	+	+	-	-
LDE3	+	+	-	-	-	+	+	+	-	-
LDE4	-	-	-	-	-	-	-	-	-	-
LDE5	+	+	-	-	-	-	+	-	-	+
LDE6	+	+	-	-	-	+	+	-	-	-
LDE7	-	-	-	-	-	-	+	+	-	-
LDE8	-	-	-	-	-	-	-	-	-	-
LDE9	-	-	-	-	-	-	-	-	-	-
LDE10	-	-	-	-	-	+	+	-	-	-
LDE11	-	-	-	-	-	-	-	-	-	+
DVI1	+	+	-	-	-	+	+	-	-	-
DVI2	+	+	-	-	-	+	+	-	-	-
DVI3	-	-	-	-	-	-	-	-	-	-
DVI4	-	-	-	-	-	-	-	-	-	-
DVI5	-	-	-	-	-	-	-	-	-	+
DVI6	-	-	-	-	-	-	-	-	-	+
DVI7	-	-	-	-	-	-	-	-	-	-
DVI8	+	-	-	-	+	+	+	+	-	-
DVI9	+	-	-	-	+	+	+	+	-	-
DVI10	+	-	-	-	+	+	+	+	-	-
DVI11	+	-	-	-	+	+	+	+	-	-
SOR1	+	+	-	-	-	+	+	-	-	-
SOR2	+	+	-	-	+	+	+	-	-	-
SOR3	-	-	-	-	-	-	-	-	-	+
SOR4	-	-	-	-	-	-	-	-	-	+
SOR5	-	-	-	-	-	-	-	-	-	+
	CMC	RAC	CC	C3	C4	C5	C6	Glu	Gal	XG

(+) indicates degradation, (-) no degradation of the respective substrates. CMC = Carboxymethyl cellulose; RAC = regenerated amorphous cellulose; CC = Crystalline cellulose; C3 to C6 = Cellobiose to Cellohexaose; Glu = Glucomannan; Gal = Galactomannan; XG = Xyloglucan.

We were able to confirm the cellulolytic activity initially observed on CMC agar plates (CTR1 and CTR3, PCO1, PCO3, PCO4 and PCO6, LDE2, LDE3, LDE5, and LDE6,

DVI1, DVI2 and DVI8-11, SOR1 and SOR2). Each of these enzymes was able to break down CMC, regenerated amorphous cellulose (RAC) and cellulose oligomers.

Interestingly, LDE10 did not show activity against cellulosic polymers but preferentially degraded cellopentaose and cellohexaose (Fig. S3). Similarly, but with much weaker efficiency, PCO5 degraded cellohexaose (Fig. S2). Together with the plate assays, our TLC analyses clearly demonstrated that beetle-derived GH45s processed cellulosic substrates using an endo-active mechanism, which suggests that these enzymes are endo-β-1,4-glucanases. Several cellulolytically active GH45s derived from the four leaf beetle species displayed additional activity towards the hemicellulose glucomannan, for example, CTR1, PCO3 and LDE2 (Table 1; Figs. S1 to S3). LDE7 exhibited the highest enzymatic activity against glucomannan, whereas its activity against amorphous cellulose substrates could be visualized only by plate assay.

Interestingly, TLC allowed us to detect several enzymes (CTR2, PCO7, LDE5, LDE11, DVI5, DVI6 and SOR3 to SOR5) which were able to degrade xyloglucan instead of cellulose (Table 1; Figs. S1 to S5). The size of the resulting breakdown products seemed to correlate with heptamers and octamers, indicating that these proteins were endo-β-1,4-xyloglucanases; however, the actual size of the resulting breakdown products is difficult to assess because some glucose moieties that make up the backbone of xyloglucan are substituted with xylose residues. LDE5 displayed activity towards amorphous cellulose substrates (Figs. 1B and S3), and is thus, according to our data, the only example of a beetle-derived GH45 able to degrade xyloglucan as well as amorphous cellulose. Additionally, each of the 37 GH45s was tested against xylan and no activity was detected (data not shown).

The Chrysomelid-derived PCO8, LDE1, LDE4, LDE8, LDE9, DVI3, DVI4 and DVI7 did not exhibit activity towards any of the substrates used in this study. We wondered whether substitutions of catalytically important residues may have caused their apparent loss of activity. To test this hypothesis, we performed an amino acid alignment of target beetle GH45 sequences, including a fungal GH45 sequence, as a reference for which the structure has been resolved (Davies et al. 1995); then we

screened for amino acid substitutions and compared these to the reference fungal sequence (Fig. 2).

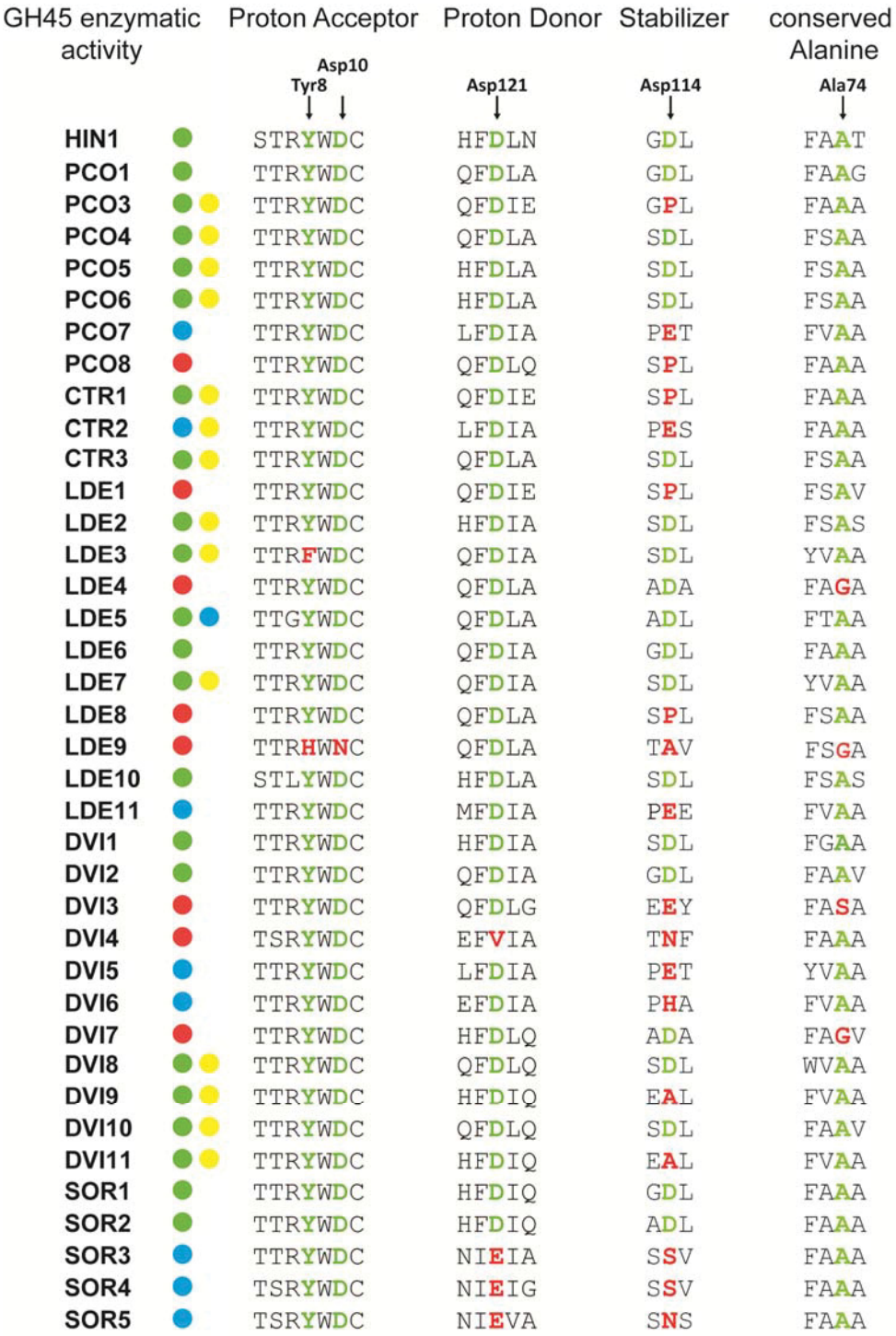


Fig.2 figure legend: see next page.

Fig.2 GH45 amino acid alignment of the catalytic residues. We used a GH45 sequence of Humicola insulens (HIN1) as reference sequence (Accession: 2ENG_A) (Davies et al. 1995). According to HIN1 we chose to investigate the catalytic residues (ASP10 and ASP121) as well as a conserved tyrosine (TYR8) of the catalytic binding site, a crucial substrate stabilizing amino acid (ASP114) and an essential conserved alanine (ALA74). Arrows indicate amino acid residue under investigation. If the respective amino acid residue is highlighted in green it is retained in comparison to the reference sequence, otherwise it is highlighted in red. GH45 enzymatic activity was color-coded based on their respective substrate specificity (Green dots = endo-β-1,4-glucanase, blue dots= endo-β-1,4-xyloglucanase, yellow dots = (gluco)mannanase, red dots= putatively non-activity).

According to Davies et al. (1995), both the proton donor (catalytic acid) and the acceptor (catalytic base) of the catalytic dyad should be aspartates (Asp10 and Asp121). In LDE9, the catalytic base was substituted for an asparagine, whereas in DVI4, the catalytic acid was substituted for a valine (Fig. 2). In both cases, the loss of the carboxyl unit of the functional group likely caused the proteins to lose the catalytic activity. No substitution event of the catalytic residues was observed for PCO8, LDE1, LDE4, LDE8, DVI3, and DVI7. Thus, we decided to investigate several conserved sites known to affect the enzymatic activity of GH45s (Davies et al. 1995). In addition, we investigated three other sites crucial for enzymatic activity: (i) a proposed stabilizing aspartate (Asp114), (ii) a conserved alanine (Ala74) and (iii) a highly conserved tyrosine (Tyr8) (Fig. 2). Apart from two substitution events of Tyr8 in LDE9 and LDE3, this amino acid remained conserved in all other beetle GH45 sequences. LDE9 already possessed a mutation in its catalytic acid, which was likely responsible for the lack of activity. In LDE3, a substitution from Tyr8 to Phe8 did not significantly impact the catalytic abilities of this protein, likely because the side-chains of both amino acids are highly similar and differ only in a single hydroxyl group. When examining the proposed stabilizing site Asp114, we observed several amino acid substitutions that correlated with a loss of activity in PCO8, DVI3, DVI4, LDE1 and LDE8. Amino acid changes at the Asp114 position were also observed in PCO3 and CTR1, but were not correlated with a loss of enzymatic activity. Since LDE4 and DVI7 appeared to have no mutation in Asp114, we screened the Ala74 residue for

substitutions; the amino acid exchange we observed, from alanine to glycine in both cases, may have caused the loss of activity in these two proteins. Altogether, amino acid substitutions at important sites could be detected in some apparently inactive GH45s, but not in all of them. It may be that the proteins for which we did not find amino acid substitutions are still active enzymes, and we have just not yet found the right substrate; alternatively, we have not yet checked all the amino acid positions, some of which could also be crucial for catalysis.

Interestingly, all Chrysomelid GH45 xyloglucanases (except LDE5 and DVI6), including *G. viridula* GVI1 from our previous study, (Busch et al. 2018a), displayed a substitution from aspartate to glutamate at the stabilizing site (Asp114) (Fig. 2). As glutamate differs from aspartate only by an additional methyl group within its side chain, we believe that this exchange may have contributed to the substrate shift. Interestingly, and in contrast to Chrysomelidae-derived GH45 xyloglucanases, we found that GH45 xyloglucanases from the Curculionidae *S. oryzae* (SOR3 to SOR5) had a substitution from aspartate to glutamate in the proton donor residue (Asp121). We also believe that, in *S. oryzae*, this particular substitution may have contributed to the preference for xyloglucan over cellulose as a substrate.

In summary, we demonstrated that each species investigated encoded at least two cellulolytic GH45s that are able to degrade amorphous cellulose. We also demonstrated that at least one GH45 per species possessed the ability to degrade only xyloglucan. Interestingly, several GH45s did not show activity on any of the substrates we tested, suggesting that they have become pseudo-enzymes or are active on substrates not tested here.

Phylogenetic analyses reveal multiple origins of GH45 genes during the evolution of Metazoa

For further insight into the evolutionary history of beetle-derived GH45 genes, we used phylogenetic analyses to reconstruct their evolutionary history. To achieve this goal, we collected amino acid sequences of GH45s available as of February 2018, including those from the CAZy database (http://www.cazy.org) (Lombard et al. 2014) as well as from several transcriptome datasets accessible at NCBI Genbank.

Interestingly, we realized that the presence of GH45 genes in arthropods was not restricted to Phytophaga beetles: these genes were also distributed in transcriptomes/genomes of species of springtails (Hexapoda: Collembola) and of species of Oribatida mites (Arthropoda: Chelicerata) (Table S2). In addition, we identified GH45 sequences in two other groups of Metazoa, namely, tardigrades and rotifers. Notably, our homology search did not retrieve any mollusk-derived GH45s; given the distant relationship of these GH45s to any of those we investigated, this absence is not surprising. The patchy distribution of GH45 genes throughout the arthropods and, more widely, the Metazoa, could be due to either the presence of GH45 genes in a common ancestor, followed by multiple gene losses, or from multiple independent acquisitions from foreign sources (i.e., HGT). To test these hypotheses, we collected a diverse set of GH45 sequences of microbial and metazoan origins resulting in 264 sequences (Table S2). Subsequently, redundancy at 90% identity level between sequences was eliminated, resulting in 201 non-redundant GH45 sequences. According to both Bayesian and maximum likelihood phylogenies (Fig. 3, S6 andS7), the arthropod-derived GH45 sequences were not monophyletic but globally fell into three separate groups. One highly supported monophyletic clade (posterior probability (PP) =1.0, bootstrap =85) grouped all the Phytophaga beetle GH45 sequences. This clade was most closely related to a group of Saccharomycetales fungi (PP=0.88, bootstrap=44). Then this clade branched to yet another group of Saccharomycetales Fungi (PP= 1.0, bootstrap=56). A second monophyletic clade grouped all the Oribatida mites GH45 sequences, although with moderate support on the branch (PP=0.72, bootstrap=3). Finally, a third clade (PP=0.96, bootstrap=14) grouped all the GH45 sequences from Collembola. This last group was not monophyletic: a bacterial GH45 sequence was interspersed within this clade and separated the Collembola GH45 sequences into two subgroups. In addition to the arthropods, the nematode GH45 sequences formed a highly supported monophyletic clade (PP= 1.0, bootstrap=84) which was connected to a clade of fungal-derived sequences. This connection was highly supported (PP= 0.93, bootstrap=69). The two other groups of Metazoa (tardigrades and rotifers) were located in a separate clade with species of Neocallimastigaceae fungi (Chytrids),

(PP=0.94, bootstrap=14). Overall, this analysis showed that neither arthropods nor, more generally, metazoan GH45 sequences, originated from a common ancestor, as they were scattered in multiple separate clades rather than forming a monophyletic metazoan clade.

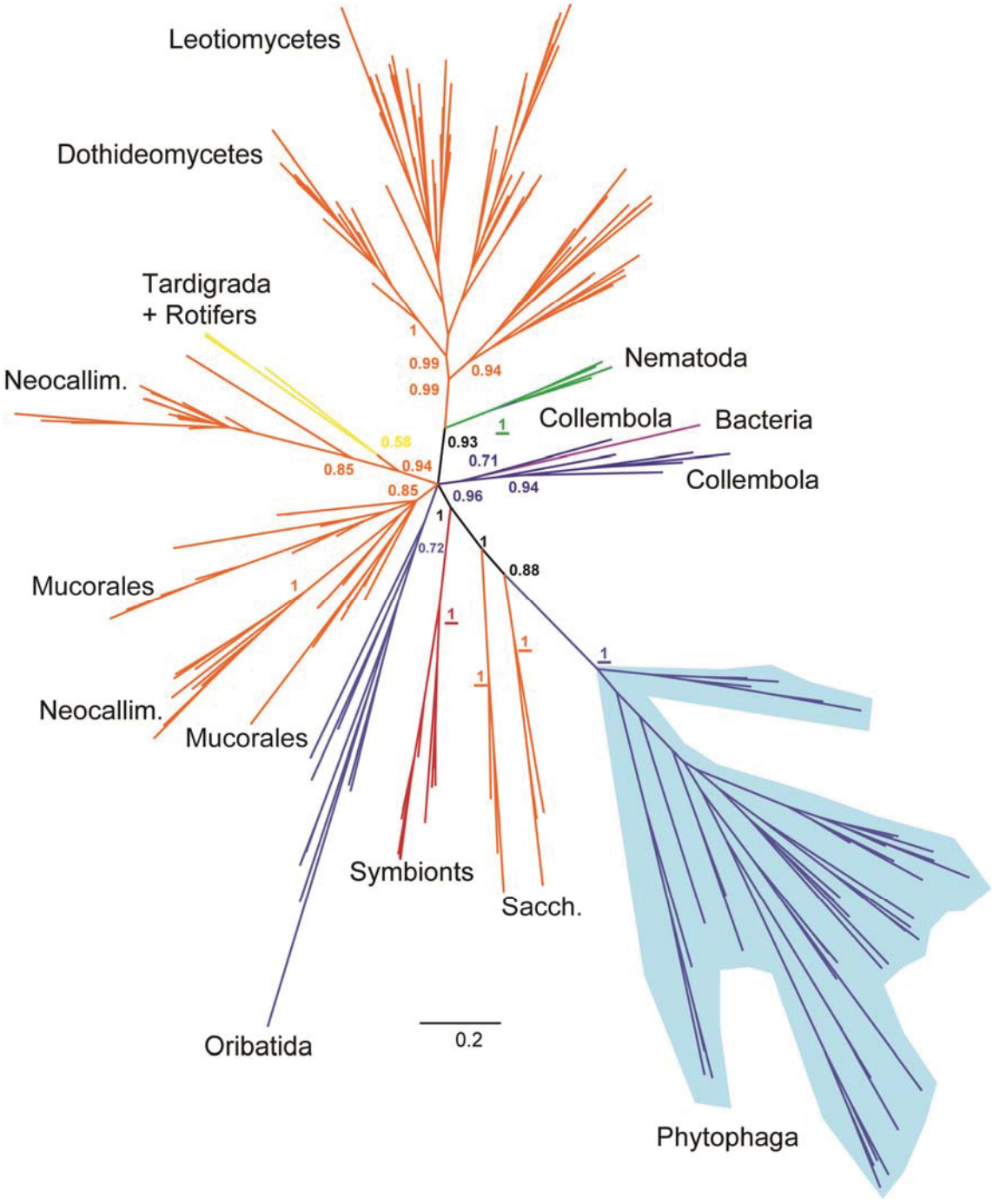


Fig. 3 figure legend: see next page

Fig. 3 Global Phylogeny encompassing GH45 proteins from various taxa. Bayesian based phylogenetic analysis of GH45s sequences. 264 GH45 sequences of microbial and metazoan origin were initially collected (see method section) and their redundancy was eliminated at 90 % sequence similarity resulting in a total of 201 sequences. Posterior probability values are given at crucial nodes. If values are depicted in bold the same branch appeared in the corresponding maximum likelihood analysis (see Fig. S6 and S7). If underlined the maximum likelihood node was highly supported (bootstrap values > 75). Detailed sequences descriptions including accession numbers are given in Table S2. Hexapoda are represented in blue, fungi in orange, protists in red, Bacteria in purple, Nematoda in green and other Metazoans in yellow. Sacch. = Saccharomycetales fungi; Neocallim. = Neocallimastigaceae fungi.

In summary, our phylogenetic analyses illustrated that the evolutionary history of GH45s in the Metazoa was complex and pointed to the possibility that this gene family evolved several times independently in multicellular organisms. More specifically, our analyses suggested that this gene family had evolved at least three times independently in arthropods. Finally, our data pointed toward an acquisition of GH45 genes by the LCA of Phytophaga beetles -- presumably through an HGT event -- from a fungal donor.

The structure of GH45 genes in Phytophaga beetles supports a single origin before the split of the Chrysomeloidea and Curculionoidea

The monophyly of the Phytophaga-derived GH45s in the above phylogenetic analyses suggests a common ancestral origin in this clade of beetles. If the presence of a GH45 in the Phytophaga beetles had resulted from a single acquisition in their LCA, we hypothesized that the GH45 genes present in current species of leaf beetles, longhorned beetles and weevils would share a common exon/intron structure. To test this hypothesis, we mined the publicly available genomes of three species of Curculionidae, including *H. hampei* (Vega et al. 2015), *D. ponderosae* (Keeling et al. 2013) and *S. oryzae* (unpublished), as well as the genomes of the Chrysomelidae *L. decemlineata* (Schoville et al. 2018) and of the Cerambycidae *A. glabripennis* (McKenna et al. 2016). We were able to retrieve the genomic sequence corresponding to each of the GH45 genes present in these beetle species, with the

exception of DPO9, which we did not find at all, and SOR3 and SOR4, which we were able to retrieve only as partial genomic sequences. Our results showed that the number of introns varied between the different species (Fig. S8). In *L. decemlineata* (representing Chrysomelidae), we identified a single intron in each of the GH45 genes (except for LDE11, which had two). For *A. glabripennis* (representing Cerambycidae), we found two introns in each of the two GH45 genes. In *H. hampei*, *D. ponderosae* and *S. oryzae* (all representing Curculionidae), the number of introns ranged from three to five. Interestingly, all GH45 genes in these five species possessed an intron placed within the part of the sequence encoding the predicted signal peptide. Apart from DPO7 and DPO8, these introns were all in phase one. This gene structure of Chrysomelid- and Curculionid-derived GH45 genes correlated well with our previous study investigating the gene structure of PCW-degrading enzymes, including GH45 genes, in the leaf beetle *Chrysomela tremula* (Pauchet et al. 2014b). The conservation of the phase and the position of this intron indicated that the LCA of the Phytophaga likely possessed a single GH45 gene having a phase one intron located in a part of the sequence encoding a putative signal peptide. To assess whether that particular intron is also present in the most closely related fungal species, we blasted the genomes of Saccharomycetaceae and Neocallimastigaceae fungi (NCBI, whole-shotgun genome database) using the protein sequence of GH45-1 of *C. tremula*. We did not detect any introns in fungal GH45 sequences (data not shown), suggesting that the proposed intron was acquired after the putative HGT event. The diversity of the overall intron-exon structure in phytophagous beetles likely resulted from subsequent and independent intron acquisition. In summary, and together with the monophyly of beetle-derived GH45s (Fig. 3), our analysis highly supports a common ancestral origin of beetle GH45.

Evolution of the GH45 family after the initial split of Chrysomeloidea and Curculionoidea

We mined publicly available transcriptome and genome datasets of Phytophaga beetles (Table S3) and collected as many GH45 sequences as possible. We curated a total of 266 GH45 sequences belonging to 42 species of Phytophaga beetles. After amino acid alignment, we decided to exclude 60 partial GH45 sequences from our

phylogenetic analysis because these were too short. We performed a “whole Phytophaga” phylogenetic analysis on the remaining 206 curated GH45 sequences using maximum likelihood (Fig. 4). Because most of the deeper nodes were poorly supported, we decided to collapse branches having a bootstrap support below 50. Our phylogenetic analysis indicated that no orthologous genes were found between species of Chrysomeloidea and Curculionoidea (Fig. 4). The only exception to this rule was found in a clade containing the xyloglucanases from the Chrysomelidae (clade m) and from the Curculionidae (clade n), which cluster together with a bootstrap support of 73 Proceeding cautiously, because the substrate switch from amorphous cellulose to xyloglucan seemed to be due to different amino acid substitutions at catalytically important sites between Chrysomelidae-derived xyloglucanases and Curculionidae-derived ones (Fig. 2), we hypothesize a single common ancestry of cellulolytic GH45s; in contrast, xyloglucanase activity likely arose through convergent evolution at list twice within the Phytophaga clade of beetles.

Clade n comprised Brentidae- and Curculionidae-derived GH45s including SOR3 to SOR5 (Fig. 4). According to our functional data, SOR3 to SOR5 act as xyloglucanases, suggesting that other GH45 proteins present in clade n may have also evolved to degrade xyloglucan. To support this hypothesis, we compared the catalytic residues of SOR3-5 to those of the other Curculionidae and Brentidae-derived GH45 sequences from clade n (Fig. S9). We detected substitutions from an aspartate to a glutamate at Asp121 in all Curculionidae-derived GH45 sequences of this clade but not in the Brentidae-derived sequences, suggesting that Curculionidae-derived GH45s of this clade were likely to possess xyloglucanase activity. Functional analyses of the Brentidae-derived sequences present in clade n will be needed to determine whether these proteins are also xyloglucanases or whether they fulfill another function.

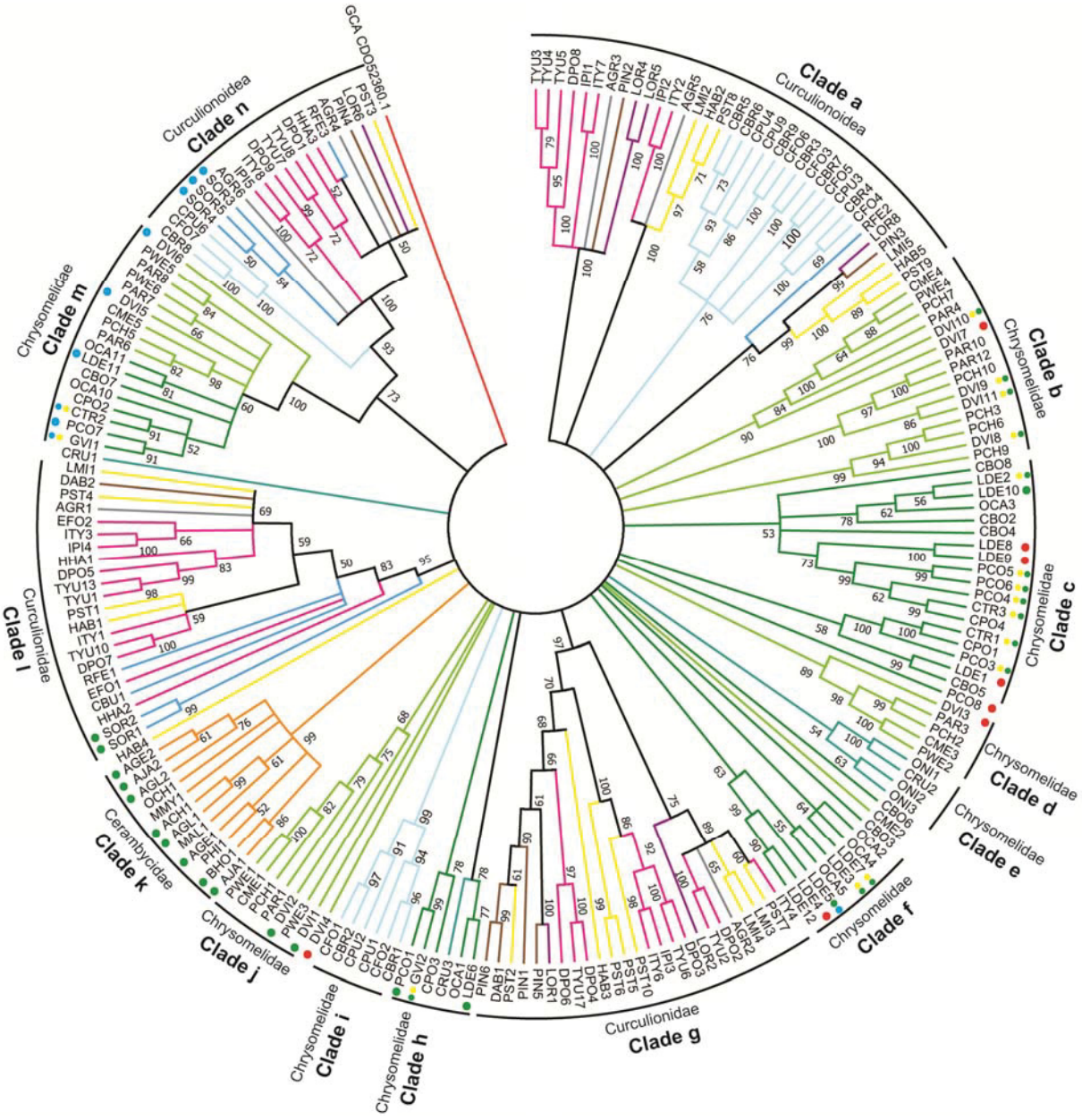


Fig. 4 figure legend: see next page

Fig. 4 Phylogenetic relationships of Phytophaga-derived GH45s. A maximum likelihood inferred phylogeny of the predicted amino acid sequences of beetle-derived GH45s. Bootstrap values are indicated at corresponding nodes. The tree was collapsed at nodes below a bootstrap value of 50. Information on sequences and their accession number are given in the Table S3. Dots indicate characterized GH45s to date and are color-coded based on their activity: green = endo-β-1,4-glucanases; blue = endo-β-1,4-xyloglucanases; yellow = (gluco)mannanases; red = no activity detected. Color coding in reference to the respective subfamily of Curculionoidea: pink = Scolytinae (Curculionidae); brown = Entiminae (Curculionidae); Purple = Cyclominae (Curculionidae); grey = Curculioninae (Curculionidae); yellow = Molytinae (Curculionidae); light blue = Brentinae (Brentidae); dark blue = Dryophthorinae (Curculionidae). Color coding in reference to the respective subfamily of Chrysomeloidea: dark green = Chrysomelinae (Chrysomelidae); light green = Galerucinae (Chrysomelidae); orange = Lamiinae (Cerambycidae); cyan = Cassidinae (Chrysomelidae). GH45 enzymatic activity was color-coded based on their respective substrate specificity (Green dots = endo-β-1,4-glucanase, blue dots= endo-β-1,4-xyloglucanase, yellow dots = (gluco)mannanase, red dots= putatively non-activity).

The second major cluster encompassed GH45s of clade I, with a bootstrap support of 95, and contained only Curculionidae-derived sequences (Fig. 4). Within this clade, we found SOR1 and SOR2, which are, according to our functional data, endo-active cellulases. Their presence in clade I implies that other GH45s of this cluster exhibit potential endo-cellulolytic activity. To further support this hypothesis, we again investigated amino acid residues of the catalytic site by comparing SOR1 and SOR2 to other GH45 sequences in this clade (Fig. S9). We did not find crucial substitutions in any of the investigated sites, implying that all GH45 proteins of this clade may have retained endo-β-1,4-glucanase activity. GH45 sequences present in the other Curculionidae and Brentidae-specific clades did not harbor any amino acid substitutions which could impair their catalytic properties, and they may all possess the ability to break down amorphous cellulose. More functional analyses will be necessary to assess the function of these proteins.

Regarding Chrysomeloidea-derived sequences, a highly supported clade (Fig. 4, clade m) contained GH45 sequences of two subfamilies of Chrysomelidae, namely, Chrysomelinae and Galerucinae. Our functional analyses revealed that this clade

contained GH45 proteins possessing xyloglucanase activity, including DVI5 and DVI6 from *D. vir. virgifera*. The remaining Galerucinae-derived GH45s, such as those from the Alticines *Phyllotreta armoraciae* and *Psylliodes chrysocephala*, present in clade m have yet to be functionally characterized. When the catalytic residues of active xyloglucanases from our study were compared to the uncharacterized GH45 sequences present in clade m, we observed that at least PAR6 and PCH5 of the Galerucinae and OCA10 and CPO2 of the Chrysomelinae had congruent substitutions (ASP114 > Glu114), which likely enabled those proteins to also degrade xyloglucan (Fig. S10). Therefore, it is highly likely that the LCA of the Chrysomelinae and the Galerucinae possessed at least two GH45 proteins, an endo-acting cellulase and a xyloglucanase.

Clade k consisted solely of Lamiinae-derived sequences and in fact encompassed all Cerambycidae-derived GH45s identified to date (Fig. 4). Several of those GH45s had been previously functionally characterized as cellulases including AJA1 and AJA2 (Pauchet et al. 2014a), AGE1 and AGE2 (Lee et al. 2004; Lee et al. 2005), AGL1 and AGL2 (McKenna et al. 2016), ACH1 (Chang et al. 2012) and BHO1 (Mei et al. 2015). Investigating their catalytic residues revealed no critical substitutions (Fig. S10), indicating that each yet-uncharacterized Cerambycidae GH45s from clade k (i.e. OCH1, MMY1, PHI and MAL1) may also possess cellulolytic activity.

In summary, our focus on Phytophaga-derived GH45s with regards to enzymatic characterization and ancestral origin allowed us to postulate that at least one GH45 protein was present in the LCA of the Phytophaga beetles and that this GH45 protein likely possessed cellulolytic activity. After the split between Chrysomeloidea and Curculionoidea, the GH45 gene family evolved through gene duplications at the family, subfamily and even genus/species level. Finally, according to our data, the ability of these beetles to break down xyloglucan, one of the major components of the primary plant cell wall, happened at least twice, once in the LCA of the Chrysomelinae and Galerucinae and once in the LCA of the Curculionidae.

Discussion

In our previous research, we found that several beetles of the Phytophaga encoded a diverse set of GH45 putative cellulases (Pauchet et al. 2010). Here we demonstrated that in each of the five Phytophaga beetles investigated, at least two of these GH45s possess cellulolytic activity. This discovery is in accordance with other previously described GH45 proteins from Insecta (Pauchet et al. 2014a), Nematoda (Kikuchi et al. 2004), Mollusca (Rahman et al. 2014), Rotifera (Szydlowski et al. 2015) and microbes (Mcgavin and Forsberg 1988).

Surprisingly, several GH45 proteins were able to degrade glucomannan in addition to cellulose. We hypothesize that GH45 bi-functionalization may have occurred as a result of the chemical similarities between cellulose and glucomannan. Glucomannan is a straight chain polymer consisting of unevenly distributed glucose and mannose moieties. GH45 cellulases could recognize two adjoining glucose moieties in the glucomannan chain, thus allowing hydrolysis to occur. Notably, enzymes specifically targeting mannans of the PCW are rare in Phytophaga beetles. So far they have been identified and characterized only in *G. viridula* and *Callosobruchus maculatus* (GH5 subfamily 10 or GH5_10) (Busch et al. 2017; Pauchet et al. 2010), and one GH5_8 has been characterized in the coffee berry borer *H. hampei* (Acuna et al. 2012a). But in contrast to the activity on glucomannan of some GH45s we observed here, those GH5_10s and GH5_8 were true mannanases, displaying activity towards galactomannan as well as glucomannan. Although our experiments suggested some GH45 cellulases were also active on glucomannan, we believe that the activity these proteins carry out could be important for the degradation of the PCW in the beetle gut. In fact, mannans, including glucomannan, can make up to 5 % of the plant primary cell wall (Scheller and Ulvskov 2010) and may be a crucial enzymatic target during PCW degradation. This hypothesis is further supported by the presence of at least one GH45 protein with some ability to degrade glucomannan in each of the Chrysomelid beetles for which we have functional data.

Another interesting discovery was that several GH45 proteins have lost their ability to use amorphous cellulose as a substrate and evolved instead to degrade xyloglucan,

the major hemicellulose of the plant primary cell wall (Pauly et al. 2013). We believe that the initial substrate shift from cellulose to xyloglucan has likely been promoted by similarities between the substrate backbones (in both cases β-1,4 linked glucose units). The major difference between cellulose and xyloglucan is that the backbone of the latter is decorated with xylose units (which in turn can be substituted by galactose and/or fucose). We presume that the substrate shift from a straight chain polysaccharide such as cellulose to a more complex one such as xyloglucan requires the similar complex adaptation of the enzyme to its novel substrate. However, in contrast to glucomannan-degrading GH45s, GH45 xyloglucanases have apparently completely lost their ability to use amorphous cellulose as a substrate. Here, we clearly demonstrated that, following several rounds of duplications, GH45s in Chrysomelid beetles have evolved novel functions in addition to their ability to break down amorphous cellulose, allowing these insects to degrade two additional major components of the PCW, namely. glucomannan and xyloglucan. This broadening of their functions further emphasizes that GH45 proteins may have likely been an important innovation during the evolution of the Phytophaga beetles and may have strongly contributed to their radiation. In summary, the ability of GH45 proteins to degrade a variety of substrates either as monospecific or as bi-functionalized enzymes indicates that these proteins are particularly prone to substrate shifts.

According to our data, the ability to break down xyloglucan using a GH45 protein has evolved at least twice independently in Phytophaga beetles, once in the LCA of Chrysomelinae and Galerucinae and once in the LCA of the Curculionidae or of the Curculionidae and Brentidae. Once the first Brentidae-derived GH45s are functionally characterized, we will know more. Given that genome/transcriptome data for a majority of families and subfamilies are lacking throughout the Phytophaga clade, we expect that other examples of independent evolution of GH45 xyloglucanases will be revealed in the future. It is important to note that the ability to degrade xyloglucan, which represents an important evolutionary innovation for Phytophaga beetles, may not be linked solely to the evolution of the GH45 family. In fact, in *A. glabripennis* (Cerambycidae: Lamiinae), a glycoside hydrolase family 5 subfamily 2 (GH5_2) protein has evolved to degrade xyloglucan; additionally, orthologous sequences of

this GH5_2 xyloglucanase have been found in other species of Lamiinae (McKenna et al. 2016).

The ability of GH45s to break down xyloglucan correlated with a substitution event from an aspartate to a glutamate residue at a stabilizing site (Asp114) within the Chrysomelidae. Interestingly, the same amino acid exchange was present in SOR3-SO5 but was located at the catalytic acid (Asp121) rather than the stabilizing site (Asp114). Aspartate and glutamate share the same functional group but differ in the length of their side chain. Thus, the preservation of the functional group coupled with an elongated side chain has likely contributed to the substrate switch of those GH45 proteins which when turned on allows xyloglucan to be degraded. Notably, DVI6 and LDE5 do not share that particular substitution but are able to degrade xyloglucan. Therefore, we believe that the transition from cellulase to xyloglucanase has not been driven solely by a single amino acid substitution, but has been triggered by changes at other positions.

Linked to these observations, our Phytophaga-focused phylogeny combined xyloglucanases of Chrysomelidae and Curculionoidea in a single well-supported clade (grouping clades m and n on Fig. 4), suggesting that the LCA of the Phytophaga may have already possessed a GH45 xyloglucanase. Despite the well-supported GH45 xyloglucanase clade, we remain skeptical about a common ancestral GH45 xyloglucanase present in the LCA of the Phytophaga based on two pieces of evidence: first, there are no GH45 xyloglucanases in cerambycid beetles, indicating that Cerambycidae have either lost their GH45 xyloglucanases or never obtained it in the first place. Given that Cerambycidae -- at least species of Lamiinae -- have evolved to degrade xyloglucan by using GH5_2, we believe that a substrate shift from cellulose to xyloglucan of a GH45 never evolved in this family of beetles, not that is was lost. Second, as described above, there are distinct substitution events in the catalytic site between proteins from species of the two superfamilies that likely led to their substrate shift. Based on these facts, we suggest that GH45 xyloglucanases have evolved convergently in both superfamilies. In contrast to GH45 xyloglucanases, GH45 cellulases were present in species of each Phytophaga family

investigated to date. This strongly suggests that a cellulolytic GH45 was present in the ancestral Phytophaga species.

According to the carbohydrate-active enzyme (CAZy) database (Lombard et al. 2014), GH45s encompass 385 sequences (as of February 2018) distributed throughout fungi, bacteria and Metazoans. Interestingly, the distribution of GH45s within Metazoans is rather patchy, encompassing to date only Nematoda (Kikuchi et al. 2004; Palomares-Rius et al. 2014), Arthropoda (Faddeeva-Vakhrusheva et al. 2016; Pauchet et al. 2010), Rotifera (Szydlowski et al. 2015) and Mollusca (Sakamoto and Toyohara 2009), and, in insects, is restricted to Phytophaga beetles. We searched several other arthropod genome/transcriptome datasets, including beetles other than Phytophaga, and all publicly available insect genomes, as well as publicly available genomes of Collembola and Oribatida mites. Except for the latter two, we were unable to retrieve GH45 sequences from arthropods. Surprisingly, our phylogenetic analyses clearly showed that the arthropod-derived GH45s, rather than clustering together, formed three separate monophyletic groups. In fact, all metazoan-derived GH45s clustered separately, forming independent monophyletic groups. The patchy distribution of GH45 sequences among Metazoa indicates either that these proteins were acquired multiple times throughout animal evolution or that massive differential gene loss occurred within multicellular organisms. The latter hypothesis appears to be less parsimonious as it implies the existence of multiple GH45s in the LCA of Ophistokonta (Fungi and Metazoa) followed by reciprocal differential gene losses and multiple independent total gene losses in many animal lineages. Intriguingly, the closest clade to the Phytophaga GH45 sequences contained fungal-derived sequences. Our phylogenetic analyses could not identify a specific donor species/group but both suggested species of Saccharomycetales or Neocallimastigaceae fungi as potential source. The most parsimonious explanation for the appearance of GH45 genes in Phytophaga beetles is that one or more genes was acquired by horizontal gene transfer (HGT) from a fungal donor. A similar scenario may have been responsible for the presence of GH45 genes in Oribatida and Collembola, but this hypothesis remains speculative until more sequences from both these orders are identified. In addition to the monophyly of Phytophaga-derived

GH45 sequences, a common origin was further suggested by the fact that the position and the phase of the first intron was (except for two cases) conserved across GH45 genes from the species of Cerambycidae, Chrysomelidae and Curculionidae for which genome data are available. If our hypotheses are correct, the LCA of all Phytophaga beetles most likely acquired a single GH45 gene from a fungal donor. As we did not find any fungal (Saccharomycetales or Neocalimastigaceae) introns corresponding to the proposed original intron, it appears that beetle-derived GH45 genes have acquired an intron after the HGT. The GH45 gene then likely underwent several duplications before the separation of the different Phytophaga clades, and these duplications continued independently after the diversification of this hyper-diverse clade of beetles.

Strikingly, nematode-derived GH45 sequences were consistently grouped together with Saccharomycetales fungi in each analysis we ran, clearly demonstrating that the closest relatives to their GH45 genes were fungal and from different fungi than the insect relatives. The origin of nematode-derived GH45 genes has been investigated, and their acquisition by HGT from a fungal source has been proposed (Kikuchi et al. 2004; Palomares-Rius et al. 2014). Here we provide the third independent confirmation of this fact.

In conclusion, our research indicated that the Phytophaga GH45s have adapted to substrate shifts. In addition to cellulose, this adaptation led to the recognition and catalysis of two additional substrates, neither of which can be enzymatically addressed by any other GH family that those insects encode. Beetles of the Chrysomelidae have evolved to break down three components of the PCW (cellulose, xyloglucan and glucomannan) by using only GH45s. In concert with GH28 pectinases encoded by each investigated species (Kirsch et al. 2014), these beetles have evolved a near-complete set of enzymatic tools with which to deconstruct the PCW, allowing them to gain access to the nutrient-rich plant cell contents; in addition, PCW-derived polysaccharides are a potential source of energy. Our data also suggest that GH45 is not an ancestral gene family but was likely acquired by the LCA of the Phytophaga from a fungal source, most likely through an HGT event. Both, enzymatic

activity and ancestral origin suggest that GH45s were likely an essential prerequisite for the adaptation allowing Phytophaga beetles to feed on plants.

Materials and methods

Production of recombinant GH45 proteins

Open reading frames (ORFs) were amplified from cDNAs using gene-specific primers based on previously described GH45 sequences of *C. tremula*, *P. cochleariae*, *L. decemlineata*, *D. virgifera virgifera* and *S. oryzae* (Pauchet et al. 2010). If necessary, full-length transcript sequences were obtained by rapid amplification of cDNA ends PCR (RACE-PCR) using RACE-ready cDNAs as described by (Pauchet et al. 2010). For downstream heterologous expression, ORFs were amplified using a forward primer designed to include a Kozak sequence and a reverse primer designed to omit the stop codon. cDNAs initially generated for the (RACE-PCR), as described by (Pauchet et al. 2010), were used as PCR template, and the PCR reactions were performed using a high-fidelity Taq polymerase (AccuPrime, Invitrogen). The PCR products were cloned into the pIB/V5-His TOPO (Invitrogen) in frame with a V5-$(His)_6$ epitope. TOP10 chemically competent *Escherichia coli* cells (Invitrogen) were transformed and incubated overnight on a LB agar plate containing ampicillin (100 µg/ml). To select constructs for which the recombinant DNA had ligated in the correct orientation, randomly picked colonies were checked by colony PCR using the OpIE2 forward primer (located on the vector) and a gene-specific reverse primer (Table S4). Positive clones were further cultured in 2x yeast extract tryptone (2xYT) medium containing 100 µg/ml ampicillin. After plasmid isolation using GeneJET Plasmid Miniprep Kit (Thermo Scientific), the recombinant plasmids were sequenced in both directions using Sanger sequencing to confirm whether the ORF has been correctly inserted into the vector. Positive constructs were then transfected in *Sf*9 insect cells (Invitrogen) using FuGENE HD (Promega) as a transfection reagent. First, successful expression was determined by transiently transfecting three clones per construct in a 24-well plate format. 72 h after transfection, the culture medium was harvested and centrifuged (16,000 x g, 5 min, 4 °C) to remove cell debris. Successful expression was verified by Western blot using the anti-V5-HRP antibody (Invitrogen). In order to collect enough material for downstream enzymatic activity assays, a single clone per construct was chosen to be transiently transfected in a 6-well plate format.

72 h after transfection, culture medium was harvested and treated as described above. The cell medium was stored at 4 °C until further use.

Enzymatic characterization

The enzymatic activity of recombinant proteins was initially tested on agarose diffusion assays using carboxymethyl cellulose (CMC) as a substrate. Agarose (1%) plates were prepared, containing 0.1 % CMC in 20 mM citrate/phosphate buffer pH 5.0. Small holes were made in the agarose matrix using cut-off pipette tips, to which 10 µl of the crude culture medium of each expressed enzyme was applied. After incubation for 16 h at 40 °C, activity was revealed by incubating the agarose plate in 0.1% Congo red for 1 h at room temperature followed by washing with 1 M NaCl until pale halos on a red background were visible. To investigate GH45 enzymatic activity in more detail, we analyzed their enzymatic breakdown products using thin layer chromatography (TLC). For that, the culture medium of transiently transfected cells was dialyzed and desalted as described in Busch et al. (2017). The following substrates were tested: CMC, Avicel, glucomannan, galactomannan and xyloglucan (all from Megazyme) with a final concentration of 0.5 %. We also tested regenerated amorphous cellulose (RAC), prepared according to Zhang et al. (2006). Additionally, we used the cello-oligomers D-(+)-biose to D-(+)-hexaose (all from Megazyme), as substrates at a final concentration of 250 ng/µl. Samples were incubated and analyzed as previously described (Busch et al. 2017). The reference standard contained 2 µg of each oligomer: glucose, cellobiose, cellotriose, cellotetraose and cellopentaose as well as isoprimeverose, xylosyl-cellobiose and the hepto-, octa- and nona saccharides of xyloglucan.

Tissue-specific gene expression

Third-instar *P. cochleariae* larvae, actively feeding on leaves of *B. rapa*, were used for total RNA extraction. Larvae were cut open from abdomen to head, and the complete gut was removed and stored separately from the rest of the body. Dissection and storage were carried out in RL solution (Analytik Jena). Three biological replicates were sampled, each containing three larvae. Total RNA was isolated using the innuPREP RNA Mini Kit (Analytik Jena), following the manufacturer's protocol. The

resulting RNA samples were subjected to DNase digestion (Ambion), and their quality was assessed using the RNA 6000 Nano LabChip kit on a 2100 Bioanalyser (both Agilent Technologies). Total RNA was used as a template to synthesize cDNAs, using the Verso cDNA synthesis kit (Thermo Scientific). The resulting cDNA samples were then used for real-time quantitative PCR (qPCR) experiments, which were performed in 96-well hard-shell PCR plates on the CFX Connect Real-Time System (both Bio-Rad). All reactions were carried out using the 2-Step QPCR SYBR Kit (Thermo Scientific), following the manufacturer's instructions. Primers were designed using the program Primer3 (version 0.4.0) (Table S4). The specific amplification of each transcript was verified by dissociation curve analysis. A standard curve for each primer pair was determined in the CFX Manager (version 3.1) based on Cq-values (quantitation cycle) of qPCRs run with a dilution series of cDNA pools. The efficiency and amplification factors of each qPCR, based on the slope of the standard curve, were calculated using an integrated efficiency calculator of the CFX manager software (version 3.1). Ribosomal protein S3 (RPS3), extracted from our *P. cochleariae* larval gut transcriptome (Kirsch et al. 2012), was used as a reference gene, and the abundance of GH45 transcripts was expressed as RNA molecules per 1000 RNA molecules of RPS3. Gene expression values were ln-transformed and significant differences between gut and rest-body were analyzed in SigmaPlot Version 11.0 using paired t-tests.

Gene structure determination

Genomic sequences of GH45 encoding genes were mined from publicly available draft genomes of *L. decemlineata* (Schoville et al. 2018), *A. glabripennis* (McKenna et al. 2016), *H. hampei* (Vega et al. 2015), *D. ponderosae* (Keeling et al. 2013) and *S. oryzae* (unpublished; accession: SAMN08382431). The intron/exon structure was determined for each gene using splign (Kapustin et al. 2008), a spliced aligner.

Amino acid alignment and Phytophaga-specific phylogenies

Sequences corresponding to Phytophaga GH45 proteins described in our previous studies were combined with those mined from several NCBI databases, such as the non-redundant protein database (ncbi_nr) and the transcriptome shotgun assembly

database (ncbi_tsa) (Table S3). In addition, transcriptome datasets generated from species of Phytophaga beetles were retrieved from the short-read archive (ncbi_sra) (Table S3) and assembled using the CLC workbench program version 11.0. Reads were loaded and quality trimmed before being assembled using standard parameters. The resulting assemblies were screened for contigs matching known beetle GH45 sequences through BLAST searches. The resulting contigs were then manually curated and used for further analysis. Amino acid alignments were carried out using MUSCLE version 3.7 implemented in MEGA7 (version 7.0.26) (Kumar et al. 2016). The maximum likelihood analysis was also conducted in MEGA7. The best model of protein evolution was determined in MEGA7 using the 'find best DNA/protein models' tool. The best model was the 'Whelan and Goldman' (WAG) model, incorporating a discrete gamma distribution (shape parameter = 5) to model evolutionary rate differences among sites (+G) and a proportion of invariable sites (+I). The robustness of the analysis was tested using 1,000 bootstrap replicates.

Large phylogenetic analysis

We used the GH45 protein sequence from *Sitophilus oryzae* (ADU33247.1) as a BLASTp query against the NCBI's non-redundant protein library with an E-value threshold of $1E^{-3}$. We retrieved the 250 best blast hits (Table S2), encompassing a majority of fungal sequences as well as various hexapod sequences (including Chrysomelidae, Curculionidae, Lamiinae and Collembola (=Entomobryomorpha)). Besides fungi and hexapods, GH45 sequences from 10 nematodes, from one Tardigrade, one Rotifer and one bacterium, as well as a few uncharacterized protists from environmental samples, were among the 250 best BLAST hits. We complemented this dataset with predicted proteins from several Oribatid mites (10 sequences) and Collembola (4 sequences) retrieved from ncbi_tsa (Table S2). This resulted in a collection of 264 sequences.

The set of 264 protein sequences was scanned against the Pfam v31 library of protein domains using the pfam_scan script with default parameters. All these proteins had one GH45 (Glyco_hydro_45 PF02015) domain on at least 98% of the expected lengths. Most of the fungal proteins possessed an ancillary carbohydrate-

binding module (either CBM1 or CBM10), but this domain was not found in other species, except the bdelloid rotifer *Adineta ricciae*. We eliminated redundancy at 90% identity level between the 264 protein sequences; we used the CD-HIT Suite server (Huang et al. 2010) and reduced the dataset to 201 non-redundant sequences, while maintaining the diversity of clades.

The non-redundant sequences were aligned using MAFFT v7.271 (Katoh and Standley 2013) with the "—auto" option to allow us to automatically select the most appropriate alignment strategy. We used trimal (Capella-Gutierrez et al. 2009) to automatically discard columns that contained more than 50% of gaps in the alignment (-gt 0.5 option) and the maximum likelihood and the Bayesian methods to reconstruct phylogenetic trees. Maximum likelihood trees were reconstructed by RAxML version 8.2.9 (Stamatakis 2014) with an estimated gamma distribution of rates of evolution across sites and an automatic selection of the fittest evolutionary model (PROTGAMMAAUTO). Bootstrap replicates were automatically stopped upon convergence (-autoMRE). Bayesian trees were reconstructed by MrBayes version 3.2.6 (Ronquist et al. 2012) with an automatic estimation of the gamma distribution of rates of evolution across sites and a mixture of evolutionary models. The number of mcmc generations was stopped once the average deviation of split frequencies was below 0.05. Twenty-five percent of the trees were burnt for calculation of the consensus tree and statistics.

Acknowledgments

We are grateful to Bianca Wurlitzer and Domenica Schnabelrauch for technical support. We thank Emily Wheeler, Boston, for editorial assistance. We express our gratitude to Franziska Beran (MPI for Chemical Ecology) for sharing the transcriptomes of *Phyllotreta armoraciae* and *Psylliodes chrysocephala* prior to publication. We are also thankful to Roy Kirsch and David G. Heckel for their input on experimental design and for fruitful discussions. This work was supported by the Max Planck Society

Supplementary Material

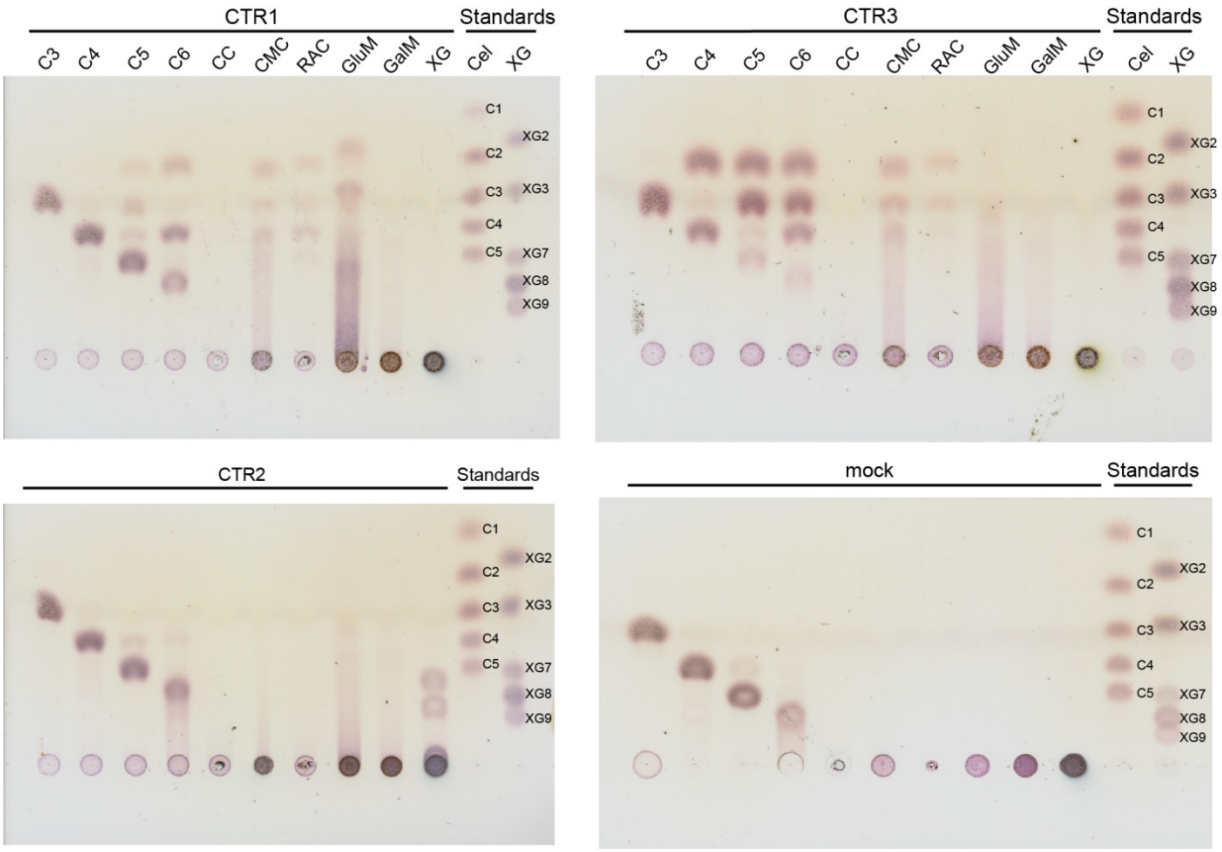


Fig. S1 Thin-layer chromatography of *C. tremula* GH45s assayed against several plant cell wall polysaccharides. Recombinant GH45s were incubated for 16 h at 40 °C with various plant polysaccharides. Their breakdown products were analyzed on TLC and visualized using 0.2 % orcinol in methane/sulphoric acid (9:1) under continuous heating. Each TLC represents an individually tested GH45 (Ctr1 to Ctr3). All GH45s were assayed against the same set of substrates, namely, cellotriose to cellohexaose (C3-C6); crystalline cellulose = avicel (CC); carboxymethyl cellulose (CMC); regenerated amorphous cellulose (RAC); glucomannan (GluM); galactomannan (GalM); xyloglucan (XG); standards: C1 = glucose, C2 – C5 = cellobiose to pentaose; XG2- XG9 = xyloglucan-oligomers.

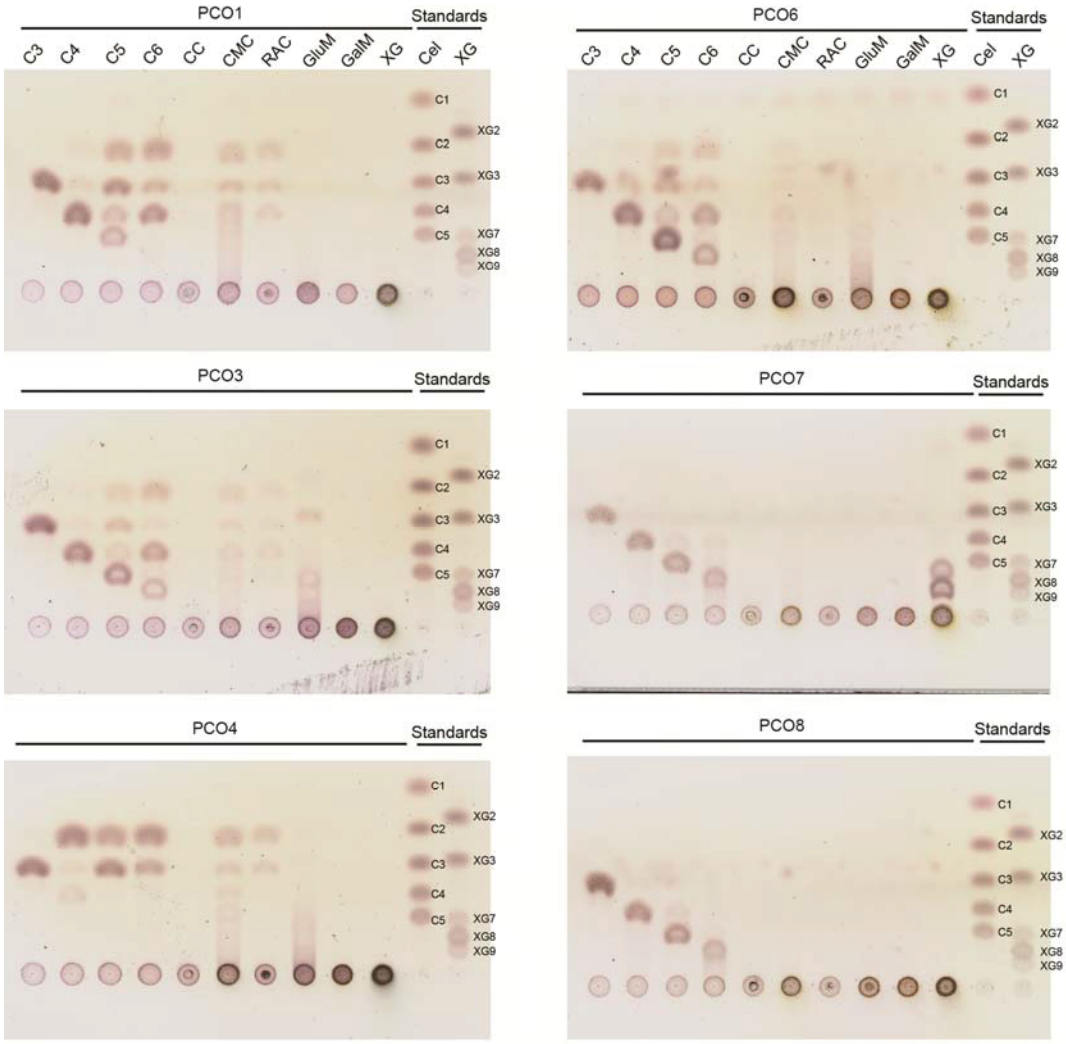


Fig. S2 Thin layer chromatography of P. cochleariae GH45s assayed against several plant cell wall polysaccharides. Recombinant GH45s were incubated for 16 h at 40 °C with various plantpolysaccharides. Their breakdown products were analyzed on TLC and visualized using 0.2 % orcinol in methane/sulphoric acid (9:1) under continuous heating. Each TLC represents an individually tested GH45 (Pco1 to Pco8). All GH45s were assayed against the same set of substrates which included cellotriose to cellohexaose (C3-C6); crystalline cellulose = avicel (CC); carboxymethyl cellulose (CMC); regenerated amorphous cellulose (RAC); glucomannan (GluM); galactomannan (GalM); xyloglucan (XG); standards: C1 = glucose, C2 – C5 = cellobiose – pentaose; XG2- XG9 = xyloglucan- oligomers.Fig. S1 Thin-layer chromatography of *C. tremula* GH45s assayed against several plant cell wall polysaccharides

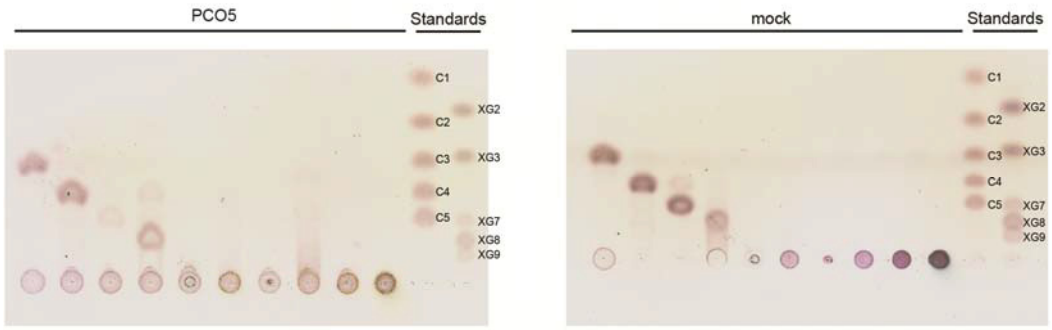


Fig. S2 continued

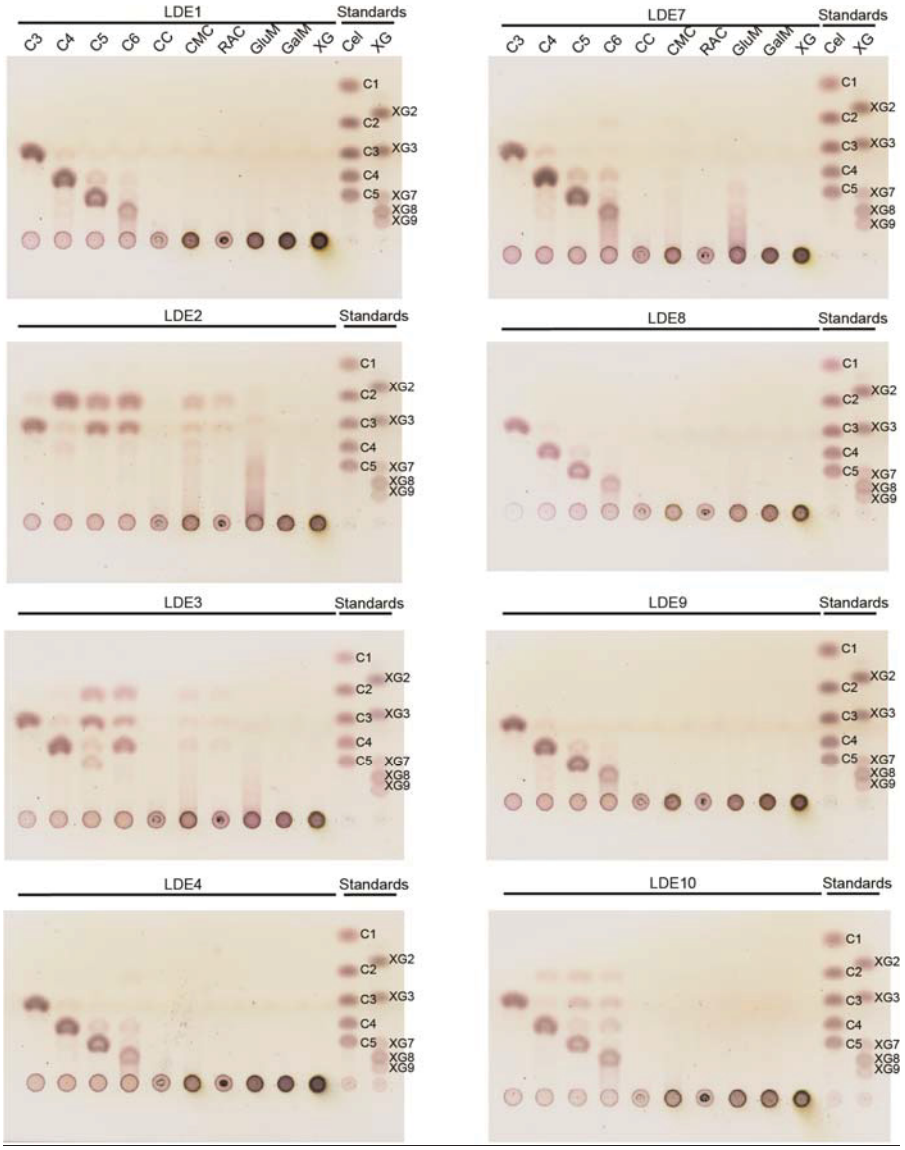


Fig. S3 Thin layer chromatography of *L. decemlineata* GH45s assayed against several plant cell wall polysaccharides. Recombinant GH45s were incubated for 16 h at 40 °C with various plant polysaccharides. Their breakdown products were analyzed on TLC and visualized using 0.2 % orcinol in methane/sulphoric acid (9:1) under continuous heating. Each TLC represents an individually tested GH45 (Lde1 to Lde11). All GH45s were assayed against the same set of substrates which included: cellotriose to cellohexaose (C3-C6); crystalline cellulose = avicel (CC); carboxymethyl cellulose (CMC); regenerated amorphous cellulose (RAC); glucomannan (GluM); galactomannan (GalM); xyloglucan (XG); standards: C1 = slucose, C2 – C5 = cellobiose – pentaose; XG2- XG9 = xyloglucan- oligomers.

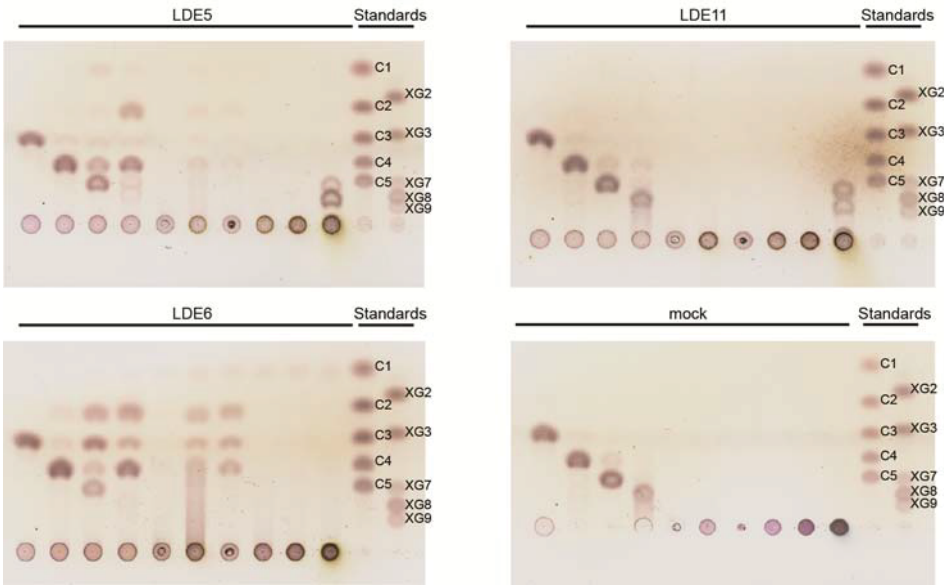


Fig. S3 continued

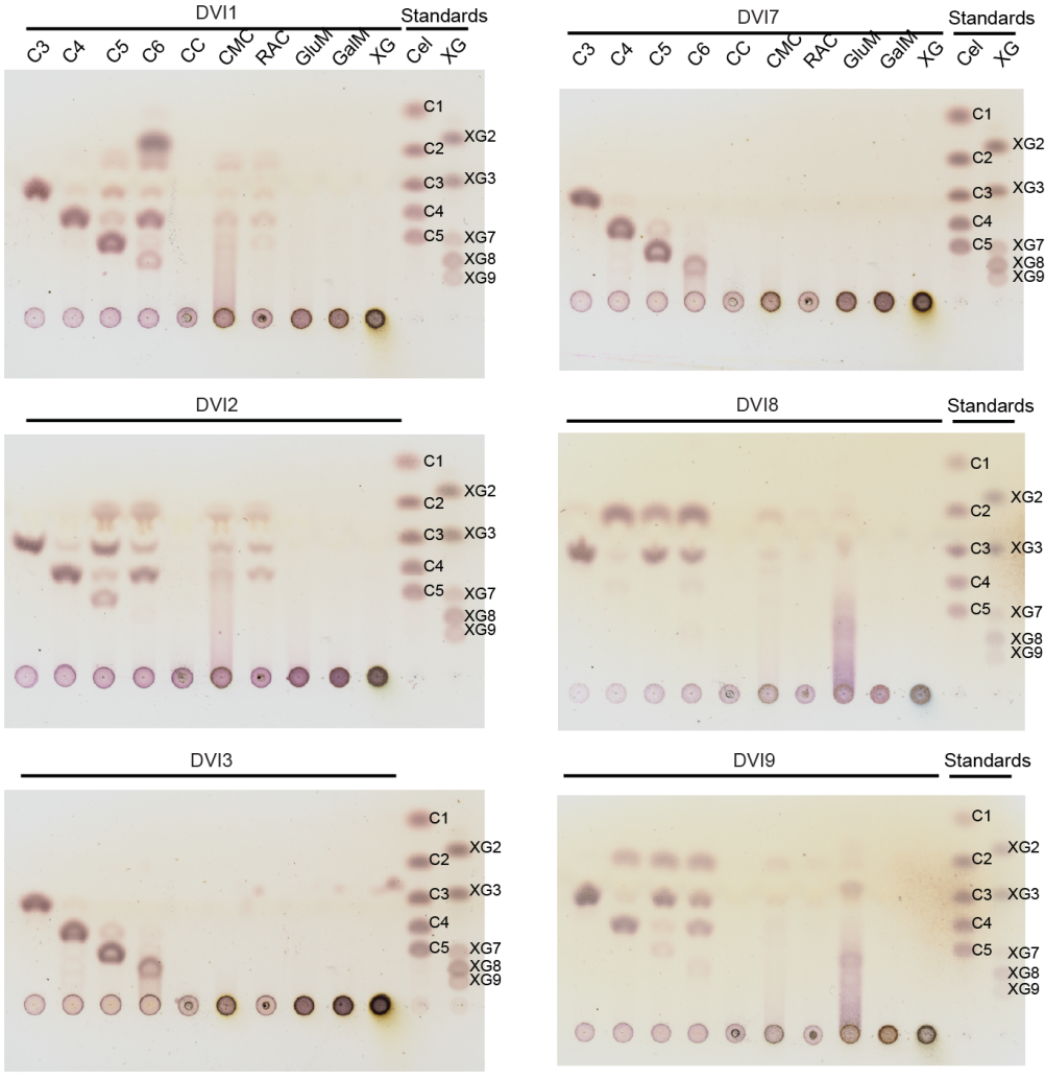


Fig. S4 Thin-layer chromatography of *D. virgifera* GH45s assayed against several plant cell wall polysaccharides. Recombinant GH45s were incubated for 16 h at 40 °C with various plant polysaccharides. Their breakdown products were analyzed on TLC and visualized using 0.2 % orcinol in methane/sulphoric acid (9:1) under continuous heating. Each TLC represents an individually tested GH45 (Dvi1 to Dvi11). All GH45s were assayed against the same set of substrates: cellotriose to cellohexaose (C3-C6); crystalline cellulose = avicel (CC); carboxymethyl cellulose (CMC); regenerated amorphous cellulose (RAC); glucomannan (GluM); galactomannan (GalM); xyloglucan (XG); standards: C1 = Glucose, C2 – C5 = cellobiose – pentaose; XG2- XG9 = xyloglucan- oligomers.

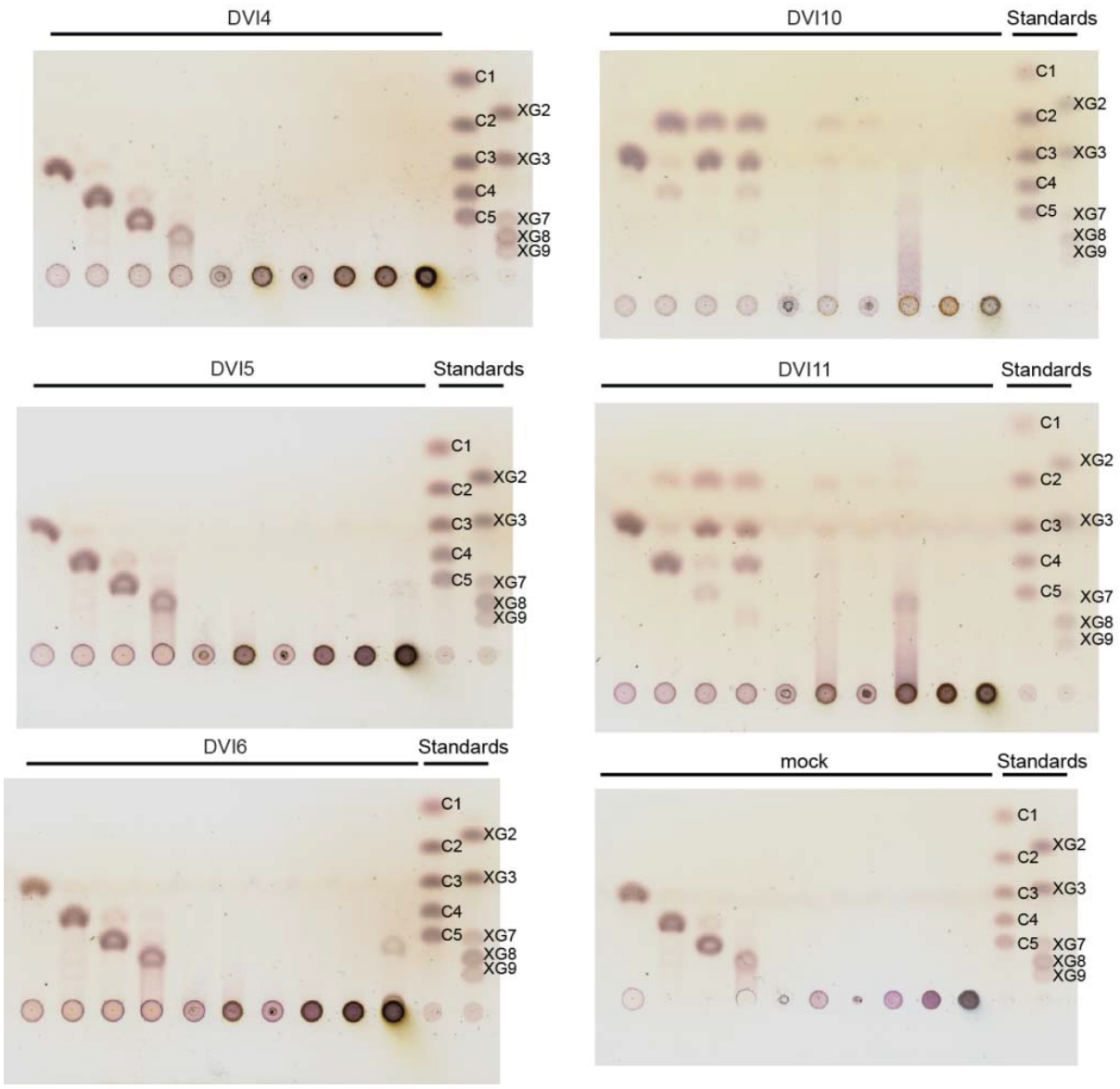


Fig. S4 continued

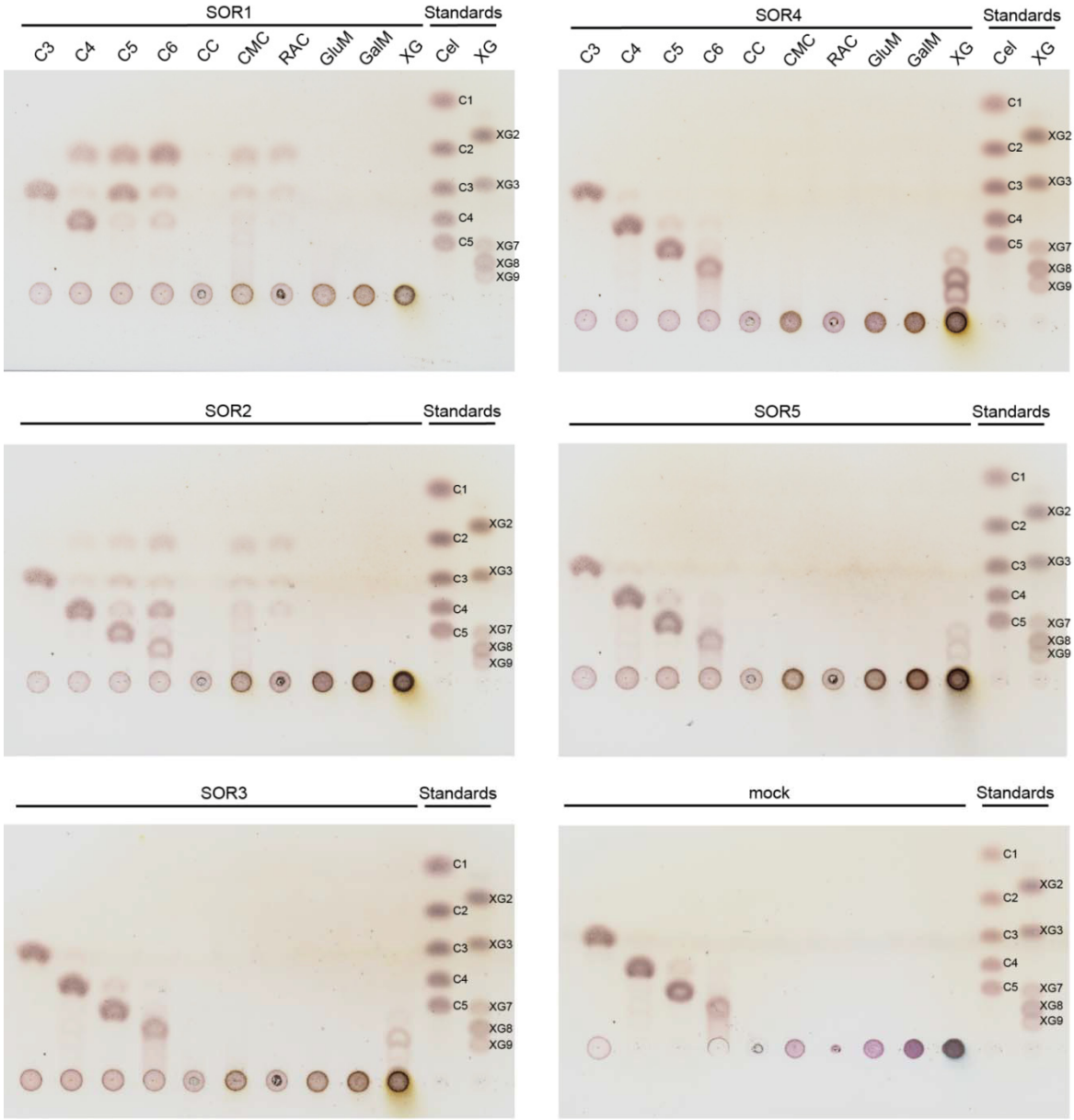


Fig. S5 Thin-layer chromatography of *S. oryzae* GH45s assayed against several plant cell wall polysaccharides. Recombinant GH45s were incubated for 16 h at 40 °C with various plant polysaccharides. Their breakdown products were analyzed on TLC and visualized using 0.2 % orcinol in methane/sulphoric acid (9:1) under continuous heating. Each TLC represents an individually tested GH45 (Sor1 toSor5). All GH45s were assayed against the same set of substrates: cellotriose to cellohexaose (C3-C6); crystalline cellulose = avicel (CC); carboxymethyl cellulose (CMC); regenerated amorphous cellulose (RAC); glucomannan (GluM); galactomannan (GalM); xyloglucan (XG); standards: C1 = glucose, C2 – C5 = cellobiose – pentaose; XG2- XG9 = xyloglucan-oligomers.

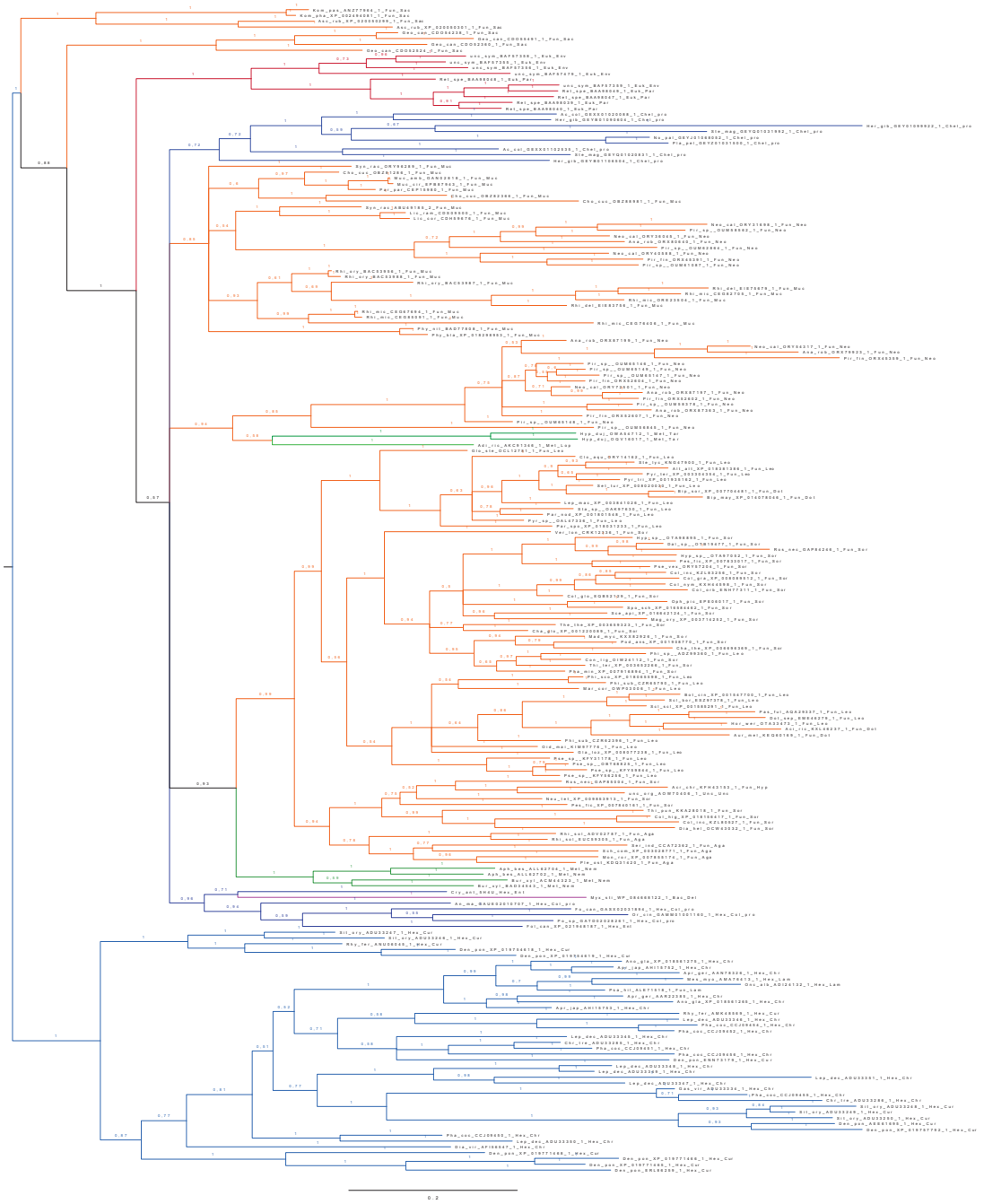
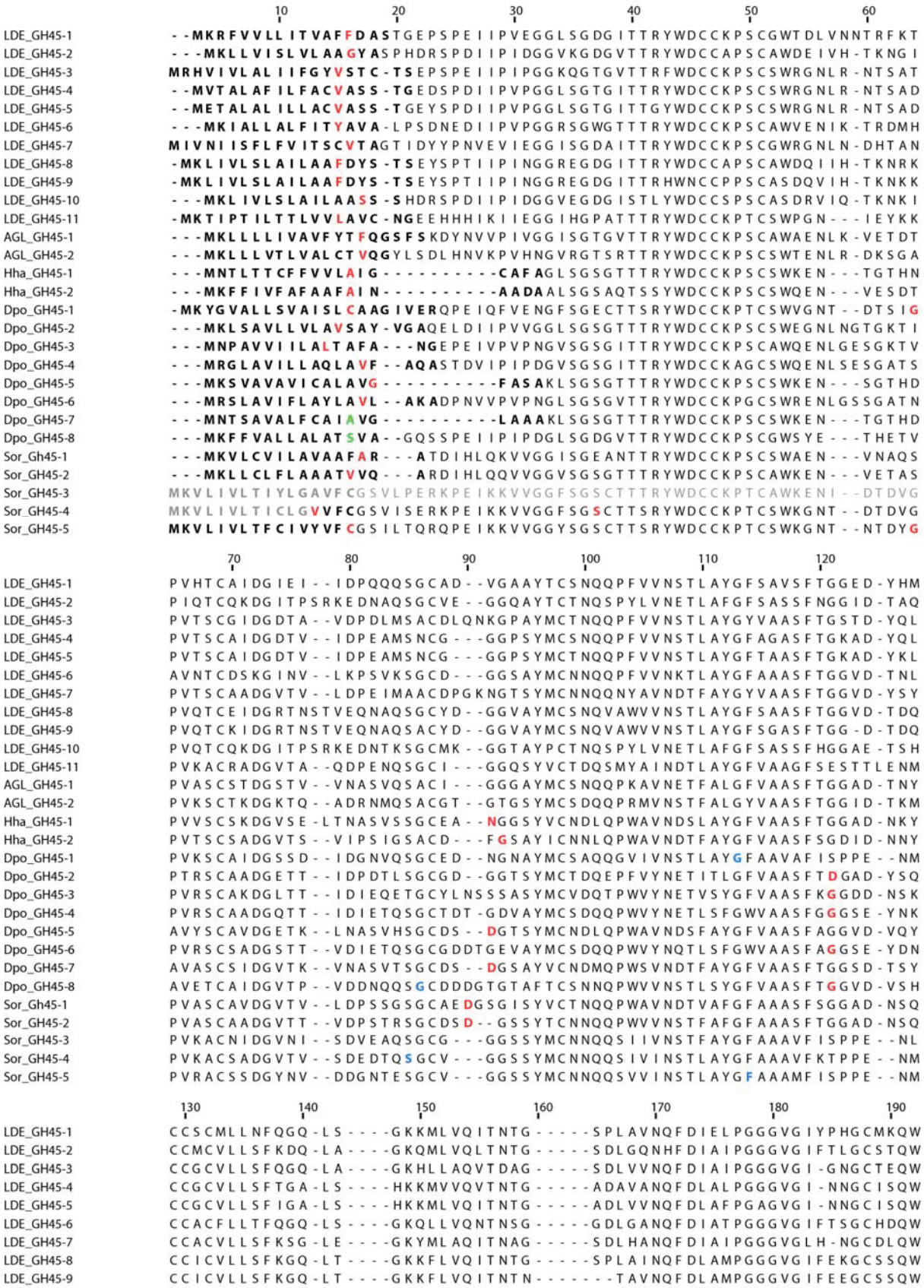


Fig. S6 Bayesian global phylogenetic analysis encompassing GH45 proteins from various taxa (expanded version of Fig. 4). 264 GH45 sequences of microbial and metazoan origin were initially collected (see Methods) and their redundancy was eliminated at 90 % sequence similarity, resulting in a total of 201 sequences. Sequence details are given in Table S2. Fungal branches are marked in orange, symbiotic protists in red, Collembola, Oribatida and Entognatha in dark blue, Coleoptera in light blue, Nematoda and Tardigrada in dark green, Rotifera in light green and bacteria in purple.

Fig. S7 Maximum likelihood inferred phylogenetic analysis encompassing GH45 proteins from various taxa. 264 GH45 sequences of microbial and metazoan origin were initially collected (see Methods), and their redundancy was eliminated at 90 % sequence similarity, resulting in a total of 201 sequences. Sequence details are given in Table S2. Fungal branches are marked in orange, symbiotic protists in red, Oribatida in dark blue, Collembola, Entognatha and Coleoptera in light blue, Nematoda in light green, Tardigrada and Rotifera in dark green and bacteria in purple.


LDE_GH45-1
LDE_GH45-2
LDE_GH45-3
LDE_GH45-4
LDE_GH45-5
LDE_GH45-6
LDE_GH45-7
LDE_GH45-8
LDE_GH45-9
LDE_GH45-10
LDE_GH45-11
AGL_GH45-1
AGL_GH45-2
Hha_GH45-1
Hha_GH45-2
Dpo_GH45-1
Dpo_GH45-2
Dpo_GH45-3
Dpo_GH45-4
Dpo_GH45-5
Dpo_GH45-6
Dpo_GH45-7
Dpo_GH45-8
Sor_Gh45-1
Sor_GH45-2
Sor_GH45-3
Sor_GH45-4
Sor_GH45-5

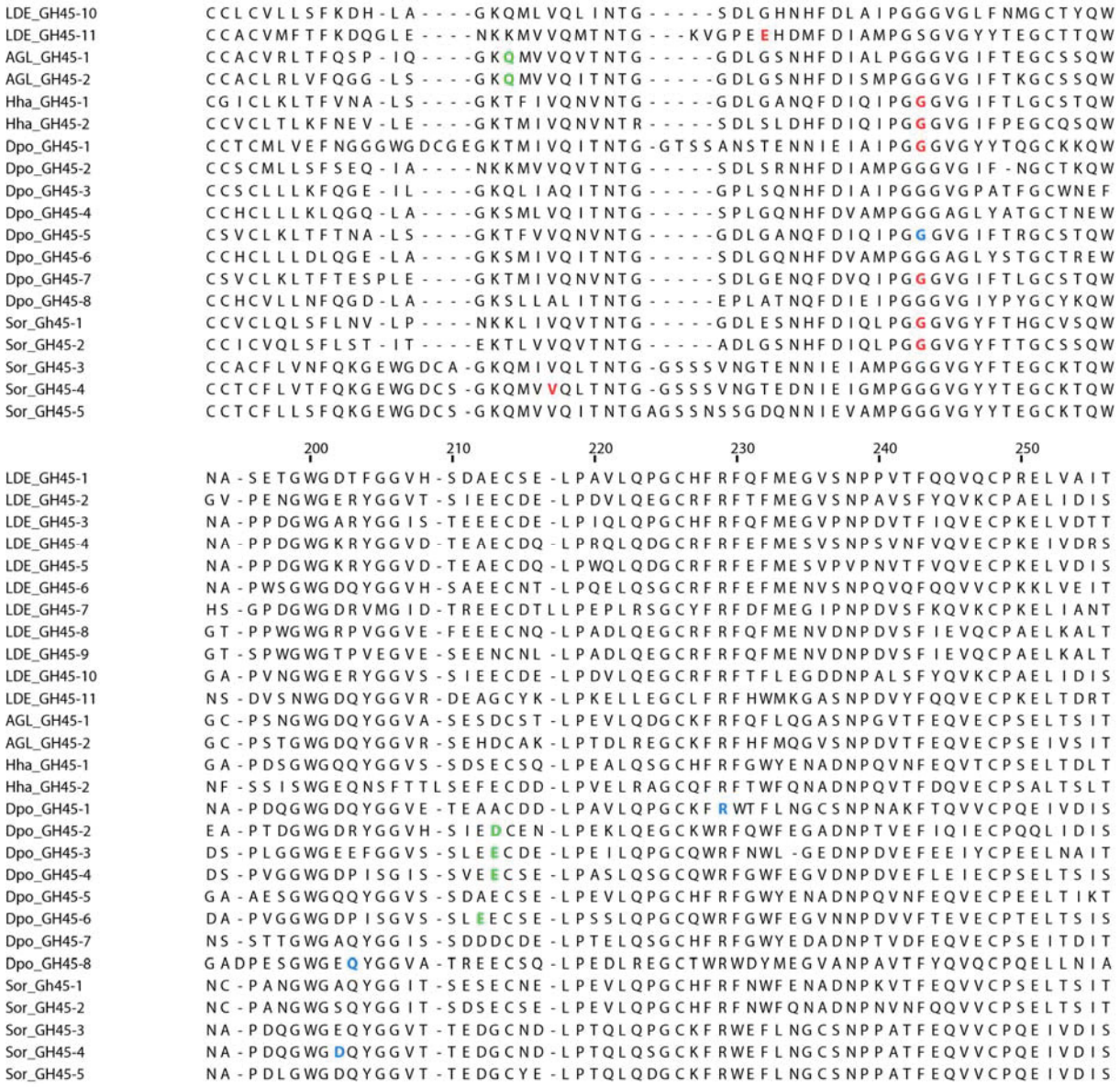


Fig. S8 Conservation of intron position in phytophagous beetles with known GH45 genome structure. Amino acid alignment of GH45 sequences derived from four different phytophagous beetles using MUSCLE. Genomic sequence information was retrieved from the genome assemblies of *L. decemlineata* (Schoville et al. 2018), *H. hampei* (Vega et al. 2015), *A. glabripennis* (McKenna et al. 2016) and *D. ponderosae* (Keeling et al. 2013). The predicted signal peptide is marked in bold letters. Intron positions are highlighted by colored amino acids according to their phase. Phase 0: green; Phase 1: red; Phase 2: blue.

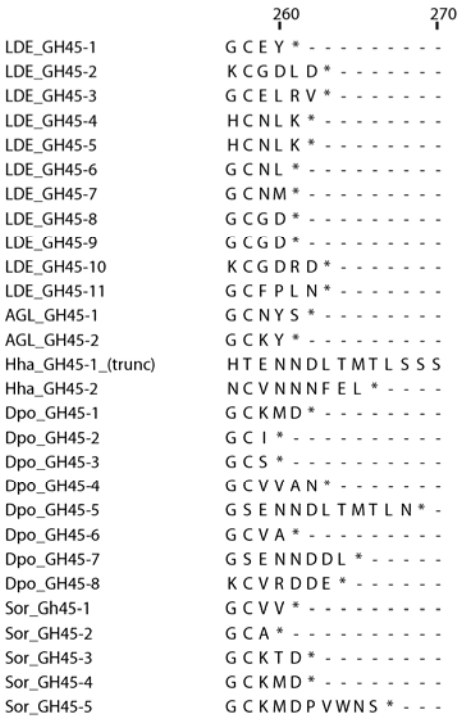


Fig. S8 continued

Fig. S9 figure legend: see next page.

Fig. S9 Amino acid alignment of the GH45 catalytic residues based on our Curculionoidea-based phylogeny. We used a GH45 sequence of Humicola insulens (HIN1) as a reference sequence (Accession: 2ENG_A) (Davies, et al. 1995). According to HIN1, we chose to investigate the catalytic residues (ASP10 and ASP121) as well as a conserved tyrosine (TYR8) of the catalytic binding site, a crucial substrate stabilizing amino acid (ASP114) and an essential conserved alanine (ALA74). Arrows indicate amino acid residue under investigation. If highlighted in green, the residue remained unchanged in comparison to HIN1; elsewise it is highlighted in red. GH45 enzymatic activity was color-coded based on the respective substrate specificity (green dots = endo-β-1,4-glucanase, blue dots = endo-β-1,4-xyloglucanase, red dots= no activity). Color coding in reference to the respective subfamily: pink = Scolytinae (Curculionidae); brown = Entiminae (Curculionidae); purple = Cyclominae (Curculionidae); gray = Curculioninae (Curculionidae); yellow = Molytinae (Curculionidae); light blue = Brentinae (Brentidae); dark blue = Dryophthorinae (Curculionidae). Each clade corresponds to the clades depicted in Fig. 4.

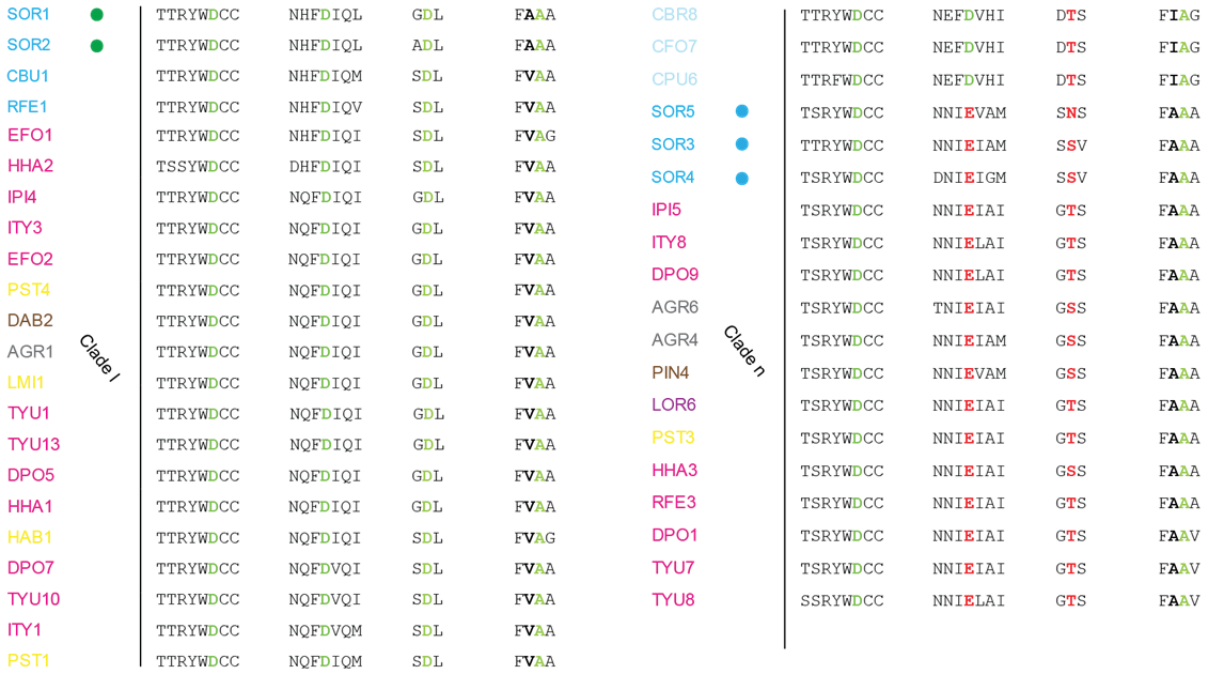


Fig. S9 continued

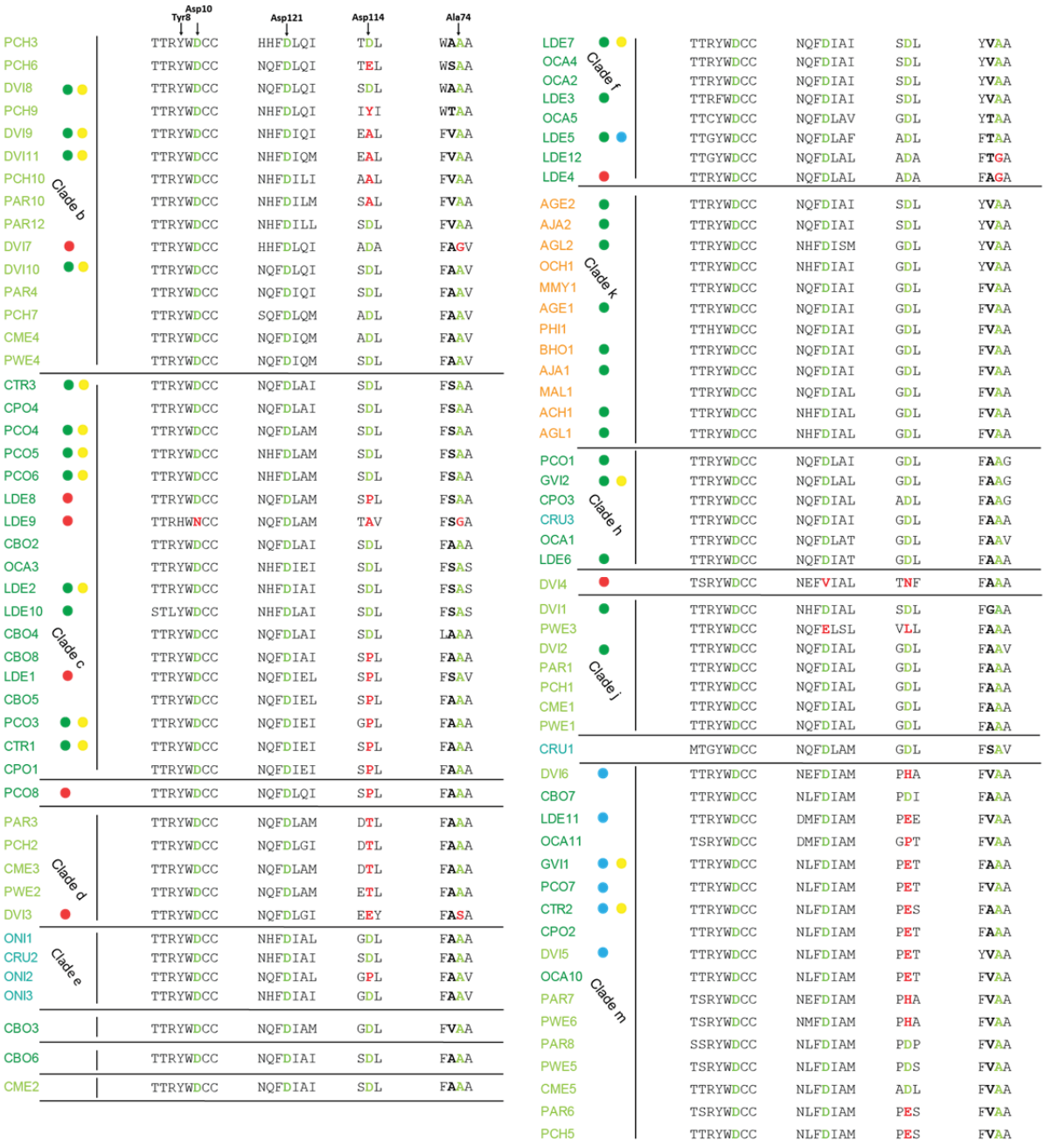


Fig. S10 figure legend: see next page.

Fig. S10 Amino acid alignment of the GH45 catalytic residues using our Chrysomeloidea-based phylogeny. We used a GH45 sequence of Humicola insulens (HIN1) as a reference sequence (Accession: 2ENG_A) (Davies et al. 1995). According to HIN1, we chose to investigate the catalytic residues (ASP10 and ASP121) as well as a conserved tyrosine (TYR8) of the catalytic binding site, a crucial substrate stabilizing amino acid (ASP114) and an essential conserved alanine (ALA74). Arrows indicate amino acid residue under investigation. If highlighted in green, the residue remained unchanged in comparison to HIN1, otherwise it is highlighted in red. GH45 enzymatic activity was color-coded based on the respective substrate specificity (green dots = endo-β-1,4-glucanase, blue dots = endo-β-1,4-xyloglucanase, red dots = no activity). Color-coding was performed with reference to the respective subfamily: dark green = Chrysomelinae (Chrysomelidae); light green = Galerucinae (Chrysomelidae); orange = Lamiinae (Cerambycidae); cyan = Cassidinae (Chrysomelidae). Each clade corresponds to the clades depicted in Fig. 4.

Table S1. Details on the beetle-derived GH45 proteins that were expressed in *Sf*9 cells.

Acronym	Species	Family	Subfamily	Accession	Reference
CTR1	*Chrysomela tremula*	Chrysomelidae	Chrysomelinae	ADU33285.1	(Pauchet et al. 2010)
CTR2	*Chrysomela tremula*	Chrysomelidae	Chrysomelinae	ADU33286.1	(Pauchet et al. 2010)
CTR3	*Chrysomela tremula*	Chrysomelidae	Chrysomelinae	MH892457	This study
LDE1	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33345.1	(Pauchet et al. 2010)
LDE2	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33346.1	(Pauchet et al. 2010)
LDE3	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33347.1	(Pauchet et al. 2010)
LDE4	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33348.1	(Pauchet et al. 2010)
LDE5	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33349.1	(Pauchet et al. 2010)
LDE6	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33350.1	(Pauchet et al. 2010)
LDE7	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	ADU33351.1	(Pauchet et al. 2010)
LDE8	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	MH892458	This study
LDE9	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	MH892459	This study
LDE10	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	MH892460	This study
LDE11	*Leptinotarsa decemlineata*	Chrysomelidae	Chrysomelinae	MH892461	This study
PCO1	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962202.1	(Kirsch et al. 2012)
PCO3	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962203.1	(Kirsch et al. 2012)
PCO4	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962204.1	(Kirsch et al. 2012)
PCO5	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962205.1	(Kirsch et al. 2012)
PCO6	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962206.1	(Kirsch et al. 2012)
PCO7	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962207.1	(Kirsch et al. 2012)
PCO8	*Phaedon cochleariae*	Chrysomelidae	Chrysomelinae	HE962208.1	(Kirsch et al. 2012)
DVI1	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	AFI56547.1	(Valencia et al. 2013)
DVI2	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892463	(Eyun et al. 2014)
DVI3	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892464	(Eyun et al. 2014)
DVI4	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892465	(Eyun et al. 2014)
DVI5	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892466	(Eyun et al. 2014)
DVI6	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892467	(Eyun et al. 2014)
DVI7	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892468	(Eyun et al. 2014)
DVI8	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892469	(Eyun et al. 2014)
DVI9	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892470	(Eyun et al. 2014)
DVI10	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892471	(Eyun et al. 2014)
DVI11	*Diabrotica vir. virgifera*	Chrysomelidae	Galerucinae	MH892472	(Eyun et al. 2014)
SOR1	*Sitophilus oryzae*	Curculionidae	Dryophtorinae	ADU33246.1	(Pauchet et al. 2010)
SOR2	*Sitophilus oryzae*	Curculionidae	Dryophtorinae	ADU33247.1	(Pauchet et al. 2010)
SOR3	*Sitophilus oryzae*	Curculionidae	Dryophtorinae	ADU33248.1	(Pauchet et al.

SOR4	*Sitophilus oryzae*	Curculionidae	Dryophtorinae	ADU33249.1	2010) (Pauchet et al. 2010)
SOR5	*Sitophilus oryzae*	Curculionidae	Dryophtorinae	ADU33250.1	(Pauchet et al. 2010)

Table S2. 250 best blast hits obtained from NCBI

Hit	Accession	Description	Species	species short	Taxo	Taxo-2	Max Score	Total Score	Query Cov	E-value	Identity	Accession
1	ADU33247.1	endo-beta-1,4-glucanase	Sitophilus oryzae	Sit.ory	Hexapod	Curculionidae	473	473	100%	9.00E-169	100%	Sit.ory-ADU33247.1-Hex-Cur
2	ADU33246.1	endo-beta-1,4-glucanase	Sitophilus oryzae	Sit.ory	Hexapod	Curculionidae	380	380	99%	3.00E-132	77%	Sit.ory-ADU33246.1-Hex-Cur
3	ANU06045.1	endo-beta-1,4-glucanase	Rhynchophorus ferrugineus	Rhy.fer	Hexapod	Curculionidae	337	337	100%	3.00E-115	69%	Rhy.fer-ANU06045.1-Hex-Cur
4	XP_018561275.1	PREDICTED: endoglucanase-like	Anoplophora glabripennis	Ano.gla	Hexapod	Chrysomeloidea	306	306	99%	1.00E-102	65%	Ano.gla-XP_018561275.1-Hex-Chr
5	AFN89565.1	cellulase 45A	Anoplophora chinensis	Ano.chi	Hexapod	Chrysomeloidea	300	300	99%	2.00E-100	64%	Ano.chi-AFN89565.1-Hex-Chr
6	AMA76413.1	cellulase 45	Mesosa myops	Mes.myo	Hexapod	Chrysomeloidea	295	295	99%	1.00E-98	62%	Mes.myo-AMA76413.1-Hex-Chr
7	XP_019754618.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	295	295	87%	3.00E-98	66%	Den.pon-XP_019754618.1-Hex-Cur
8	XP_019754619.1	PREDICTED: putative endoglucanase type	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	292	292	91%	2.00E-97	63%	Den.pon-XP_019754619.1-Hex-Cur
9	XP_019754620.1	PREDICTED: putative endoglucanase type	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	292	292	91%	3.00E-97	63%	Den.pon-XP_019754620.1-Hex-Cur
10	ALE71518.1	cellulase	Psacothea hilaris	Psa.hil	Hexapod	Chrysomeloidea	287	287	99%	4.00E-95	60%	Psa.hil-ALE71518.1-Hex-Chr
11	AAR22385.1	cellulase II	Apriona germari	Apr.ger	Hexapod	Chrysomeloidea	285	285	99%	2.00E-94	56%	Apr.ger-AAR22385.1-Hex-Chr
12	AHI15753.1	glycoside hydrolase family 45	Apriona japonica	Apr.jap	Hexapod	Chrysomeloidea	282	282	93%	3.00E-93	61%	Apr.jap-AHI15753.1-Hex-Chr
13	AKH90729.1	cellulase I	Batocera horsfieldi	Bat.hor	Hexapod	Chrysomeloidea	280	280	99%	1.00E-92	60%	Bat.hor-AKH90729.1-Hex-Chr
14	AAU44973.1	cellulase I	Apriona germari	Apr.ger	Hexapod	Chrysomeloidea	280	280	99%	1.00E-92	60%	Apr.ger-AAU44973.1-Hex-Chr
15	AHI15752.1	glycoside hydrolase family 45	Apriona japonica	Apr.jap	Hexapod	Chrysomeloidea	280	280	99%	3.00E-92	60%	Apr.jap-AHI15752.1-Hex-Chr
16	AFI56547.1	endogenous endoglucanase	Diabrotica virgifera virgifera	Dia.vir	Hexapod	Chrysomeloidea	278	278	100%	1.00E-91	56%	Dia.vir-AFI56547.1-Hex-Chr
17	XP_018561265.1	PREDICTED: endoglucanase-like	Anoplophora glabripennis	Ano.gla	Hexapod	Chrysomeloidea	275	275	100%	1.00E-90	55%	Ano.gla-XP_018561265.1-Hex-Chr
18	XP_019771468.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	275	275	99%	1.00E-90	57%	Den.pon-XP_019771468.1-Hex-Cur
19	CCJ09450.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	273	273	90%	7.00E-90	63%	Pha.coc-CCJ09450.1-Hex-Chr
20	ADI24132.1	endo-beta-1,4-glucanase precursor	Oncideres albomarginata chan	Onc.alb	Hexapod	Chrysomeloidea	273	273	99%	1.00E-89	56%	Onc.alb-ADI24132.1-Hex-Chr
21	AMK48569.1	glycosyl hydrolase family 45	Rhynchophorus ferrugineus	Rhy.fer	Hexapod	Curculionidae	270	270	99%	2.00E-88	55%	Rhy.fer-AMK48569.1-Hex-Cur
22	AAN78326.1	cellulase	Apriona germari	Apr.ger	Hexapod	Chrysomeloidea	268	268	99%	1.00E-87	58%	Apr.ger-AAN78326.1-Hex-Chr
23	ADU33350.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	267	267	99%	2.00E-87	56%	Lep.dec-ADU33350.1-Hex-Chr
24	ADU33345.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	264	264	89%	3.00E-86	57%	Lep.dec-ADU33345.1-Hex-Chr
25	ADU33346.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	263	263	100%	9.00E-86	53%	Lep.dec-ADU33346.1-Hex-Chr
26	XP_019771466.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	262	262	98%	3.00E-85	52%	Den.pon-XP_019771466.1-Hex-Cur
27	CCJ09454.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	258	258	100%	1.00E-83	54%	Pha.coc-CCJ09454.1-Hex-Chr
28	ORY96289.1	RlpA-like double-psi beta-barrel-protein	Syncephalastrum racemosum	Syn.rac	Fungi	Mucorales	262	262	89%	2.00E-83	60%	Syn.rac-ORY96289.1-Fun-Muc
29	CCJ09452.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	256	256	100%	4.00E-83	53%	Pha.coc-CCJ09452.1-Hex-Chr
30	OBZ81286.1	Endoglucanase-5, partial	Choanephora cucurbitarum	Cho.cuc	Fungi	Mucorales	255	255	90%	4.00E-83	58%	Cho.cuc-OBZ81286.1-Fun-Muc
31	ADU33285.1	endo-beta-1,4-glucanase	Chrysomela tremula	Chr.tre	Hexapod	Chrysomeloidea	256	256	88%	5.00E-83	58%	Chr.tre-ADU33285.1-Hex-Chr
32	CCJ09451.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	255	255	99%	1.00E-82	54%	Pha.coc-CCJ09451.1-Hex-Chr
33	ERL86259.1	hypothetical protein D910_03669	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	255	255	100%	1.00E-82	52%	Den.pon-ERL86259.1-Hex-Cur
34	OBZ82366.1	Endoglucanase-5	Choanephora cucurbitarum	Cho.cuc	Fungi	Mucorales	258	258	90%	2.00E-82	58%	Cho.cuc-OBZ82366.1-Fun-Muc
35	XP_019771465.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	254	254	98%	4.00E-82	52%	Den.pon-XP_019771465.1-Hex-Cur
36	AEE61846.1	unknown	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	253	253	98%	8.00E-82	51%	Den.pon-AEE61846.1-Hex-Cur
37	XP_019772608.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	253	253	100%	1.00E-81	51%	Den.pon-XP_019772608.1-Hex-Cur
38	ORE01467.1	glycoside hydrolase	Rhizopus microsporus var. mic	Rhi.mic	Fungi	Mucorales	251	251	89%	2.00E-81	57%	Rhi.mic-ORE01467.1-Fun-Muc
39	ADU33348.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	252	252	97%	2.00E-81	54%	Lep.dec-ADU33348.1-Hex-Chr
40	BAF57358.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	251	251	96%	2.00E-81	53%	unc.sym-BAF57358.1-Euk-Env
41	ABU49185.2	endoglucanase	Syncephalastrum racemosum	Syn.rac	Fungi	Mucorales	254	254	90%	6.00E-81	56%	Syn.rac-ABU49185.2-Fun-Muc
42	CCJ09453.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	251	251	90%	6.00E-81	55%	Pha.coc-CCJ09453.1-Hex-Chr
43	EIE78620.1	hypothetical protein RO3G_03324	Rhizopus delemar RA 99-880	Rhi.del	Fungi	Mucorales	253	253	90%	9.00E-81	57%	Rhi.del-EIE78620.1-Fun-Muc
44	BAF57355.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	250	250	96%	1.00E-80	54%	unc.sym-BAF57355.1-Euk-Env
45	BAC53956.1	endo-beta-1,4-D-glucanase	Rhizopus oryzae	Rhi.ory	Fungi	Mucorales	253	253	90%	1.00E-80	57%	Rhi.ory-BAC53956.1-Fun-Muc
46	BAC53987.1	endo-glucanase RCE2	Rhizopus oryzae	Rhi.ory	Fungi	Mucorales	254	254	91%	2.00E-80	55%	Rhi.ory-BAC53987.1-Fun-Muc
47	ADU33334.1	endo-beta-1,4-glucanase	Gastrophysa viridula	Gas.vir	Hexapod	Chrysomeloidea	249	249	97%	2.00E-80	53%	Gas.vir-ADU33334.1-Hex-Chr
48	O97401.1	RecName: Full=Endoglucanase; AltName	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	249	249	91%	4.00E-80	58%	Pha.coc-O97401.1-Hex-Chr
49	BAC53988.1	endo-glucanase RCE3	Rhizopus oryzae	Rhi.ory	Fungi	Mucorales	252	252	90%	7.00E-80	57%	Rhi.ory-BAC53988.1-Fun-Muc
50	CEG67694.1	Putative Endo-glucanase RCE3	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	251	251	89%	8.00E-80	57%	Rhi.mic-CEG67694.1-Fun-Muc

51	CEI91022.1	Putative Endo-glucanase RCE3	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	251	251	89%	9.00E-80	57% Rhi.mic-CEI91022.1-Fun-Muc
52	CEG85091.1	Putative Endo-glucanase RCE2	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	251	251	89%	2.00E-79	57% Rhi.mic-CEG85091.1-Fun-Muc
53	CEJ05250.1	hypothetical protein LRAMOSA10860	Lichtheimia ramosa	Lic.ram	Fungi	Mucorales	253	251	91%	2.00E-79	54% Lic.ram-CDS09500.1-Fun-Muc
54	CDS09500.1	partial	Lichtheimia corymbifera JMRC	Lic.cor	Fungi	Mucorales	252	253	91%	5.00E-79	54% Lic.cor-CDH59676.1-Fun-Muc
55	CDH59676.1	unknown	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	243	252	91%	9.00E-78	51% Den.pon-AEE61893.1-Hex-Cur
56	AEE61893.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	243	243	100%	1.00E-77	50% Pha.coc-CCJ09456.1-Hex-Chr
57	CCJ09456.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	243	243	99%	1.00E-77	50% Pha.coc-CCJ09456.1-Hex-Chr
58	CDO54238.1	conserved hypothetical protein. Putative	Geotrichum candidum	Geo.can	Fungi	Saccharomycetales	250	250	92%	2.00E-77	54% Geo.can-CDO54238.1-Fun-Sac
59	ORX87199.1	glycoside hydrolase, partial	Anaeromyces robustus	Ana.rob	Fungi	Neocallimastigaceae	242	242	86%	3.00E-77	57% Ana.rob-ORX87199.1-Fun-Neo
60	ANZ77964.1	BA75_04680T0	Komagataella pastoris	Kom.pas	Fungi	Saccharomycetales	254	254	91%	3.00E-77	58% Kom.pas-ANZ77964.1-Fun-Sac
61	ADU33349.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	241	241	98%	5.00E-77	52% Lep.dec-ADU33349.1-Hex-Chr
62	ALL62703.1	beta-1,4-endoglucanase 2	Aphelenchoides besseyi	Aph.bes	Meta	Nematode	240	240	89%	6.00E-77	55% Aph.bes-ALL62703.1-Met-Nem
63	OCL12781.1	glycoside hydrolase family 45 protein	Glonium stellatum	Glo.ste	Fungi	Leotiomycetes	240	240	99%	7.00E-77	49% Glo.ste-OCL12781.1-Fun-Leo
64	OUM65148.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	239	239	90%	7.00E-77	54% Pir.sp.-OUM65148.1-Fun-Neo
65	EIE75679.1	hypothetical protein RO3G_00383	Rhizopus delemar RA 99-880	Rhi.del	Fungi	Mucorales	240	240	97%	9.00E-77	49% Rhi.del-EIE75679.1-Fun-Muc
66	BAF57479.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	239	239	98%	1.00E-76	53% unc.sym-BAF57479.1-Euk-Env
67	XP_019766961.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	239	239	99%	2.00E-76	50% Den.pon-XP_019766961.1-Hex-Cur
68	CCJ09455.1	glycoside hydrolase family 45 protein	Phaedon cochleariae	Pha.coc	Hexapod	Chrysomeloidea	239	239	97%	2.00E-76	51% Pha.coc-CCJ09455.1-Hex-Chr
69	XP_021945337.1	endoglucanase-5-like	Folsomia candida	Fol.can	Hexapod	Collembola	238	238	99%	5.00E-76	50% Fol.can-XP_021945337.1-Hex-Col
70	EIE83756.1	hypothetical protein RO3G_08461	Rhizopus delemar RA 99-880	Rhi.del	Fungi	Mucorales	239	239	99%	7.00E-76	49% Rhi.del-EIE83756.1-Fun-Muc
71	CDO55491.1	conserved hypothetical protein. Putative	Geotrichum candidum	Geo.can	Fungi	Saccharomycetales	243	243	96%	1.00E-75	51% Geo.can-CDO55491.1-Fun-Sac
72	CRK12336.1	hypothetical protein BN1708_010432	Verticillium longisporum	Ver.lon	Fungi	Sordariomycetes	237	237	87%	1.00E-75	55% Ver.lon-CRK12336.1-Fun-Sor
73	OUM65146.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	237	237	87%	1.00E-75	55% Pir.sp.-OUM65146.1-Fun-Neo
74	OUM58378.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	238	238	89%	1.00E-75	53% Pir.sp.-OUM58378.1-Fun-Neo
75	ADU33351.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	237	237	91%	2.00E-75	53% Lep.dec-ADU33351.1-Hex-Chr
76	XP_002494081.1	Hypothetical protein PAS_chr4_0643	Komagataella phaffii GS115	Kom.pha	Fungi	Saccharomycetales	248	248	91%	3.00E-75	59% Kom.pha-XP_002494081.1-Fun-Sac
77	CDO52360.1	Conserved hypothetical protein. Putative	Geotrichum candidum	Geo.can	Fungi	Saccharomycetales	238	238	95%	9.00E-75	53% Geo.can-CDO52360.1-Fun-Sac
78	BAA98048.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	234	234	91%	9.00E-75	54% Ret.spe-BAA98048.1-Euk-Par
79	XP_018065598.1	hypothetical protein LY89DRAFT_540480	Phialocephala scopiformis	Phi.sco	Fungi	Leotiomycetes	236	236	91%	9.00E-75	50% Phi.sco-XP_018065598.1-Fun-Leo
80	ALL62704.1	beta-1,4-endoglucanase 3	Aphelenchoides besseyi	Aph.bes	Meta	Nematode	234	234	89%	1.00E-74	53% Aph.bes-ALL62704.1-Met-Nem
81	ORY36046.1	endoglucanase-5, partial	Neocallimastix californiae	Neo.cal	Fungi	Neocallimastigaceae	234	234	89%	1.00E-74	56% Neo.cal-ORY36046.1-Fun-Neo
82	ORY72501.1	endo-beta-D-1,4-glucanase	Neocallimastix californiae	Neo.cal	Fungi	Neocallimastigaceae	238	238	90%	1.00E-74	54% Neo.cal-ORY72501.1-Fun-Neo
83	OTA98895.1	glycoside hydrolase family 45 protein	Hypoxylon sp. CI-4A	Hyp.sp.	Fungi	Sordariomycetes	234	234	100%	2.00E-74	50% Hyp.sp.-OTA98895.1-Fun-Sor
84	XP_020050299.1	glycoside hydrolase family 45 protein	Ascoidea rubescens DSM 1968	Asc.rub	Fungi	Saccharomycetales	239	239	91%	2.00E-74	54% Asc.rub-XP_020050299.1-Fun-Sac
85	CEG66157.1	Putative Endo-glucanase RCE3	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	234	234	96%	3.00E-74	49% Rhi.mic-CEG66157.1-Fun-Muc
86	CEI89328.1	Putative Endo-glucanase RCE3	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	234	234	96%	4.00E-74	49% Rhi.mic-CEI89328.1-Fun-Muc
87	OUM67787.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	232	232	86%	4.00E-74	53% Pir.sp.-OUM67787.1-Fun-Neo
88	CEG82705.1	Putative Endo-glucanase RCE2	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	233	233	99%	4.00E-74	47% Rhi.mic-CEG82705.1-Fun-Muc
89	ORE08357.1	glycoside hydrolase	Rhizopus microsporus var. mic	Rhi.mic	Fungi	Mucorales	233	233	97%	4.00E-74	48% Rhi.mic-ORE08357.1-Fun-Muc
90	OWA52637.1	Endoglucanase-5	Hypsibius dujardini	Hyp.duj	Meta	Tardigrada	236	236	86%	6.00E-74	57% Hyp.duj-OWA52637.1-Met-Tar
91	ORE08136.1	glycoside hydrolase	Rhizopus microsporus var. mic	Rhi.mic	Fungi	Mucorales	233	233	99%	6.00E-74	47% Rhi.mic-ORE08136.1-Fun-Muc
92	ACV50414.1	endo-beta-1,4-glucanase	Cryptopygus antarcticus	Cry.ant	Hexapod	Collembola	233	233	98%	6.00E-74	50% Cry.ant-ACV50414.1-Hex-Col
93	ORY14162.1	RlpA-like double-psi beta-barrel-protein	Clohesyomyces aquaticus	Clo.aqu	Fungi	Leotiomycetes	234	234	99%	6.00E-74	47% Clo.aqu-ORY14162.1-Fun-Leo
94	ORE23504.1	glycoside hydrolase	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	233	233	96%	6.00E-74	48% Rhi.mic-ORE23504.1-Fun-Muc
95	ALL62702.1	beta-1,4-endoglucanase 1	Aphelenchoides besseyi	Aph.bes	Meta	Nematode	233	233	88%	7.00E-74	55% Aph.bes-ALL62702.1-Met-Nem
96	XP_001547700.1	hypothetical protein BC1G_13862	Botrytis cinerea B05.10	Bot.cin	Fungi	Leotiomycetes	237	237	90%	7.00E-74	54% Bot.cin-XP_001547700.1-Fun-Leo
97	5H4UA	Chain A, Crystal Structure Of Cellulase Fr	Cryptopygus antarcticus	Cry.ant	Hexapod	Collembola	233	233	98%	7.00E-74	50% Cry.ant-5H4UA-Hex-Col
98	ORY31698.1	hypothetical protein LY90DRAFT_705285	Neocallimastix californiae	Neo.cal	Fungi	Neocallimastigaceae	237	237	86%	8.00E-74	57% Neo.cal-ORY31698.1-Fun-Neo
99	EMR88707.1	putative glycoside hydrolase family 45 pr	Botrytis cinerea BcDW1	Bot.cin	Fungi	Leotiomycetes	236	236	90%	8.00E-74	54% Bot.cin-EMR88707.1-Fun-Leo
100	XP_003659323.1	glycoside hydrolase family 45 protein	Thermothelomyces thermoph	The.the	Fungi	Sordariomycetes	232	232	87%	8.00E-74	54% The.the-XP_003659323.1-Fun-Sor
101	KNG47900.1	glycoside hydrolase family 45 protein	Stemphylium lycopersici	Ste.lyc	Fungi	Leotiomycetes	232	232	94%	1.00E-73	46% Ste.lyc-KNG47900.1-Fun-Leo
102	OUM62864.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	232	232	86%	1.00E-73	55% Pir.sp.-OUM62864.1-Fun-Neo
103	OTB19477.1	glycoside hydrolase family 45 protein	Daldinia sp. EC12	Dal.sp.	Fungi	Sordariomycetes	232	232	88%	2.00E-73	51% Dal.sp.-OTB19477.1-Fun-Sor
104	OUM58562.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	238	238	91%	2.00E-73	55% Pir.sp.-OUM58562.1-Fun-Neo
105	BAD95808.1	endo-beta-D-1,4-glucanase	Mucor circinelloides	Muc.cir	Fungi	Mucorales	235	235	85%	2.00E-73	54% Muc.cir-BAD95808.1-Fun-Muc
106	GAN02818.1	glycoside hydrolase family 45 protein	Mucor ambiguus	Muc.amb	Fungi	Mucorales	235	235	85%	2.00E-73	54% Muc.amb-GAN02818.1-Fun-Muc

107	CEG66246.1	Putative Endo-glucanase RCE3	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	231	231	99%	2.00E-73	47% Rhi.mic-CEG66246.1-Fun-Muc
108	CEI02254.1	Putative Endo-glucanase RCE3	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	231	231	99%	3.00E-73	46% Rhi.mic-CEI02254.1-Fun-Muc
109	ADU33248.1	endo-beta-1,4-glucanase	Sitophilus oryzae	Sit.ory	Hexapod	Curculionidae	231	231	99%	3.00E-73	49% Sit.ory-ADU33248.1-Hex-Cur
110	XP_008020030.1	glycoside hydrolase family 45 protein	Setosphaeria turcica Et28A	Set.tur	Fungi	Leotiomycetes	231	231	86%	3.00E-73	51% Set.tur-XP_008020030.1-Fun-Leo
111	OUM65149.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	232	232	90%	4.00E-73	52% Pir.sp.-OUM65149.1-Fun-Neo
112	GAP85004.1	putative endoglucanase type K	Rosellinia necatrix	Ros.nec	Fungi	Sordariomycetes	231	231	87%	4.00E-73	55% Ros.nec-GAP85004.1-Fun-Sor
113	OTA38694.1	hypothetical protein BTJ68_01212	Hortaea werneckii EXF-2000	Hor.wer	Fungi	Leotiomycetes	236	236	99%	4.00E-73	50% Hor.wer-OTA38694.1-Fun-Leo
114	CEP15980.1	hypothetical protein	Parasitella parasitica	Par.par	Fungi	Mucorales	236	236	85%	4.00E-73	54% Par.par-CEP15980.1-Fun-Muc
115	AQA29337.1	hypothetical protein 44	Passalora fulva	Pas.ful	Fungi	Leotiomycetes	230	230	88%	5.00E-73	54% Pas.ful-AQA29337.1-Fun-Leo
116	KXH53573.1	glycosyl hydrolase family 45	Colletotrichum simmondsii	Col.sim	Fungi	Sordariomycetes	230	230	90%	6.00E-73	51% Col.sim-KXH53573.1-Fun-Sor
117	KXX82926.1	putative endoglucanase type K	Madurella mycetomatis	Mad.myc	Fungi	Sordariomycetes	231	231	97%	6.00E-73	49% Mad.myc-KXX82926.1-Fun-Sor
118	ADU33347.1	endo-beta-1,4-glucanase	Leptinotarsa decemlineata	Lep.dec	Hexapod	Chrysomeloidea	230	230	98%	8.00E-73	52% Lep.dec-ADU33347.1-Hex-Chr
119	ORY54317.1	endo-glucanase RCE3	Neocallimastix californiae	Neo.cal	Fungi	Neocallimastigaceae	234	234	86%	8.00E-73	54% Neo.cal-ORY54317.1-Fun-Neo
120	XP_021948187.1	endoglucanase-5-like	Folsomia candida	Fol.can	Hexapod	Collembola	233	233	85%	9.00E-73	55% Fol.can-XP_021948187.1-Hex-Col
121	BAD77808.1	endo-beta-D-1,4-glucanase	Phycomyces nitens	Phy.nit	Fungi	Mucorales	233	233	89%	1.00E-72	56% Phy.nit-BAD77808.1-Fun-Muc
122	ENN73179.1	hypothetical protein YQE_10233	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	244	244	100%	1.00E-72	50% Den.pon-ENN73179.1-Hex-Cur
123	XP_003841026.1	hypothetical protein LEMA_P089560.1	Leptosphaeria maculans JN3	Lep.mac	Fungi	Leotiomycetes	233	233	86%	1.00E-72	51% Lep.mac-XP_003841026.1-Fun-Leo
124	KFZ16307.1	hypothetical protein V502_05163, partial	Pseudogymnoascus sp. VKM F	Pse.sp.	Fungi	Leotiomycetes	230	230	87%	1.00E-72	52% Pse.sp.-KFZ16307.1-Fun-Leo
125	ADV02787.1	endo-1,4-beta-D-glucanase	Rhizoctonia solani	Rhi.sol	Fungi	Agaricomycetes	229	229	90%	2.00E-72	54% Rhi.sol-ADV02787.1-Fun-Aga
126	KFY46015.1	hypothetical protein V495_02698, partial	Pseudogymnoascus sp. VKM F	Pse.sp.	Fungi	Leotiomycetes	230	230	87%	2.00E-72	51% Pse.sp.-KFY46015.1-Fun-Leo
127	ORX52604.1	hypothetical protein BCR36DRAFT_58254	Piromyces finnis	Pir.fin	Fungi	Neocallimastigaceae	233	233	86%	2.00E-72	54% Pir.fin-ORX52604.1-Fun-Neo
128	OWP03006.1	hypothetical protein B2J93_3632	Marssonina coronariae	Mar.cor	Fungi	Leotiomycetes	232	232	87%	2.00E-72	53% Mar.cor-OWP03006.1-Fun-Leo
129	BAD95809.1	endo-beta-D-1,4-glucanase	Mucor circinelloides	Muc.cir	Fungi	Mucorales	234	234	85%	2.00E-72	54% Muc.cir-BAD95809.1-Fun-Muc
130	XP_007704481.1	glycoside hydrolase family 45 protein	Bipolaris sorokiniana ND90Pr	Bip.sor	Fungi	Dothideomycetes	229	229	86%	2.00E-72	50% Bip.sor-XP_007704481.1-Fun-Dot
131	XP_018381386.1	endoglucanase-5	Alternaria alternata	Alt.alt	Fungi	Leotiomycetes	228	228	96%	2.00E-72	49% Alt.alt-XP_018381386.1-Fun-Leo
132	KKA28018.1	hypothetical protein TD95_003702	Thielaviopsis punctulata	Thi.pun	Fungi	Sordariomycetes	233	233	99%	2.00E-72	52% Thi.pun-KKA28018.1-Fun-Sor
133	EPB87943.1	hypothetical protein HMPREF1544_05234	Mucor circinelloides f. circinel	Muc.cir	Fungi	Mucorales	234	234	85%	2.00E-72	54% Muc.cir-EPB87943.1-Fun-Muc
134	BAG71490.1	family 45 cellulase homologue	eukaryotic synthetic construct	euk.syn	Eukar	Unclass	228	228	86%	3.00E-72	55% euk.syn-BAG71490.1-Euk-Unc
135	BAF57359.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	228	228	86%	3.00E-72	55% unc.sym-BAF57359.1-Euk-Env
136	ORX52607.1	hypothetical protein BCR36DRAFT_58254	Piromyces finnis	Pir.fin	Fungi	Neocallimastigaceae	232	232	90%	4.00E-72	54% Pir.fin-ORX52607.1-Fun-Neo
137	BAA98041.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	228	228	86%	4.00E-72	55% Ret.spe-BAA98041.1-Euk-Par
138	ORX79923.1	hypothetical protein BCR32DRAFT_32789	Anaeromyces robustus	Ana.rob	Fungi	Neocallimastigaceae	233	233	86%	5.00E-72	55% Ana.rob-ORX79923.1-Fun-Neo
139	XP_009853913.1	hypothetical protein NEUTE1DRAFT_5025	Neurospora tetrasperma FGSC	Neu.tet	Fungi	Sordariomycetes	230	230	96%	6.00E-72	54% Neu.tet-XP_009853913.1-Fun-Sor
140	ORY40588.1	hypothetical protein LY90DRAFT_510372	Neocallimastix californiae	Neo.cal	Fungi	Neocallimastigaceae	232	232	86%	6.00E-72	53% Neo.cal-ORY40588.1-Fun-Neo
141	OAK97630.1	glycoside hydrolase	Stagonospora sp. SRC1lsM3a	Sta.sp.	Fungi	Leotiomycetes	228	228	94%	6.00E-72	48% Sta.sp.-OAK97630.1-Fun-Leo
142	BAA98036.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	227	227	86%	6.00E-72	54% Ret.spe-BAA98036.1-Euk-Par
143	BAA98039.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	227	227	86%	6.00E-72	54% Ret.spe-BAA98039.1-Euk-Par
144	GAP84246.1	putative glycoside hydrolase family 45 pr	Rosellinia necatrix	Ros.nec	Fungi	Sordariomycetes	228	228	100%	7.00E-72	47% Ros.nec-GAP84246.1-Fun-Sor
145	OUM65147.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	228	228	84%	7.00E-72	54% Pir.sp.-OUM65147.1-Fun-Neo
146	XP_018298953.1	glycoside hydrolase family 45 protein	Phycomyces blakesleeanus NR	Phy.bla	Fungi	Mucorales	231	231	89%	8.00E-72	55% Phy.bla-XP_018298953.1-Fun-Muc
147	ES297378.1	hypothetical protein SBOR_2262	Sclerotinia borealis F-4128	Scl.bor	Fungi	Leotiomycetes	231	231	90%	8.00E-72	53% Scl.bor-ES297378.1-Fun-Leo
148	KIM97776.1	carbohydrate-binding module family 1 pr	Oidiodendron maius Zn	Oid.mai	Fungi	Leotiomycetes	230	230	87%	1.00E-71	54% Oid.mai-KIM97776.1-Fun-Leo
149	KZL83256.1	glycoside hydrolase family 45 protein	Colletotrichum incanum	Col.inc	Fungi	Sordariomycetes	227	227	87%	1.00E-71	52% Col.inc-KZL83256.1-Fun-Sor
150	OWY43663.1	endoglucanase-5	Alternaria alternata	Alt.alt	Fungi	Leotiomycetes	227	227	86%	1.00E-71	52% Alt.alt-OWY43663.1-Fun-Leo
151	OTA33473.1	hypothetical protein BTJ68_05609	Hortaea werneckii EXF-2000	Hor.wer	Fungi	Leotiomycetes	232	232	99%	1.00E-71	49% Hor.wer-OTA33473.1-Fun-Leo
152	BAF57357.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	226	226	91%	1.00E-71	53% unc.sym-BAF57357.1-Euk-Env
153	KFY31178.1	hypothetical protein V493_01325, partial	Pseudogymnoascus sp. VKM F	Pse.sp.	Fungi	Leotiomycetes	230	230	87%	2.00E-71	51% Pse.sp.-KFY31178.1-Fun-Leo
154	OWA54712.1	Endoglucanase-5	Hypsibius dujardini	Hyp.duj	Meta	Tardigrada	240	240	86%	2.00E-71	56% Hyp.duj-OWA54712.1-Met-Tar
155	ORX87363.1	hypothetical protein BCR32DRAFT_32447	Anaeromyces robustus	Ana.rob	Fungi	Neocallimastigaceae	232	232	89%	2.00E-71	53% Ana.rob-ORX87363.1-Fun-Neo
156	XP_033304354.1	hypothetical protein PTT_16920	Pyrenophora teres f. teres 0-1	Pyr.ter	Fungi	Leotiomycetes	226	226	96%	2.00E-71	47% Pyr.ter-XP_033304354.1-Fun-Leo
157	XP_007714177.1	glycoside hydrolase family 45 protein	Bipolaris zeicola 26-R-13	Bip.zei	Fungi	Dothideomycetes	226	226	86%	2.00E-71	50% Bip.zei-XP_007714177.1-Fun-Dot
158	EPE06017.1	glycoside hydrolase family 45 protein	Ophiostoma piceae UAMH 113	Oph.pic	Fungi	Sordariomycetes	226	226	96%	3.00E-71	50% Oph.pic-EPE06017.1-Fun-Sor
159	BAA98049.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	226	226	87%	3.00E-71	55% Ret.spe-BAA98049.1-Euk-Par
160	ORX79922.1	hypothetical protein BCR32DRAFT_29413	Anaeromyces robustus	Ana.rob	Fungi	Neocallimastigaceae	230	230	86%	3.00E-71	54% Ana.rob-ORX79922.1-Fun-Neo
161	XP_957107.1	endoglucanase V	Neurospora crassa OR74A	Neu.cra	Fungi	Sordariomycetes	228	228	96%	3.00E-71	53% Neu.cra-XP_957107.1-Fun-Sor
162	ORX87197.1	hypothetical protein BCR32DRAFT_32455	Anaeromyces robustus	Ana.rob	Fungi	Neocallimastigaceae	231	231	86%	3.00E-71	54% Ana.rob-ORX87197.1-Fun-Neo

163	XP_007833017.1	hypothetical protein PFICI_06245	Pestalotiopsis fici W106-1	Pes.fic	Fungi	Sordariomycetes	226	226	88%	3.00E-71	50%	Pes.fic-XP_007833017.1-Fun-Sor
164	KEQ60169.1	hypothetical protein M437DRAFT_55064	Aureobasidium melanogenum	Aur.mel	Fungi	Dothideomycetes	229	229	89%	3.00E-71	51%	Aur.mel-KEQ60169.1-Fun-Dot
165	EQB52129.1	glycosyl hydrolase family 45	Colletotrichum gloeosporioide	Col.glo	Fungi	Sordariomycetes	226	226	87%	3.00E-71	51%	Col.glo-EQB52129.1-Fun-Sor
166	XP_014078046.1	glycoside hydrolase family 45 protein	Bipolaris maydis ATCC 48331	Bip.may	Fungi	Dothideomycetes	225	225	86%	4.00E-71	50%	Bip.may-XP_014078046.1-Fun-Dot
167	BAA98038.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	225	225	86%	4.00E-71	54%	Ret.spe-BAA98038.1-Euk-Par
168	KZL69663.1	glycoside hydrolase family 45 protein	Colletotrichum tofieldiae	Col.tof	Fungi	Sordariomycetes	225	225	87%	4.00E-71	52%	Col.tof-KZL69663.1-Fun-Sor
169	CZR65790.1	related to endoglucanase	Phialocephala subalpina	Phi.sub	Fungi	Leotiomycetes	228	228	86%	5.00E-71	52%	Phi.sub-CZR65790.1-Fun-Leo
170	XP_001585291.1	hypothetical protein SS1G_13860	Sclerotinia sclerotiorum 1980 U	Scl.scl	Fungi	Leotiomycetes	229	229	87%	5.00E-71	52%	Scl.scl-XP_001585291.1-Fun-Leo
171	CEL58599.1	Putative endoglucanase type K OS=Fusari	Rhizoctonia solani AG-1 IB	Rhi.sol	Fungi	Agaricomycetes	225	225	100%	6.00E-71	50%	Rhi.sol-CEL58599.1-Fun-Aga
172	OAL47336.1	glycoside hydrolase	Pyrenochaeta sp. DS3sAY3a	Pyr.sp.	Fungi	Leotiomycetes	225	225	86%	6.00E-71	49%	Pyr.sp.-OAL47336.1-Fun-Leo
173	AGY80100.1	cellobiohydrolase family protein 45	Chaetomium thermophilum	Cha.the	Fungi	Sordariomycetes	225	225	88%	6.00E-71	52%	Cha.the-AGY80100.1-Fun-Sor
174	ENH77311.1	beta-endoglucanase	Colletotrichum orbiculare MAI	Col.orb	Fungi	Sordariomycetes	225	225	87%	6.00E-71	53%	Col.orb-ENH77311.1-Fun-Sor
175	CDO52524.1	Conserved hypothetical protein. Putative	Geotrichum candidum	Geo.can	Fungi	Saccharomycetales	234	234	91%	6.00E-71	51%	Geo.can-CDO52524.1-Fun-Sac
176	ADZ99360.1	beta-1,4-endoglucanase	Phialophora sp. CGMCC 3328	Phi.sp.	Fungi	Leotiomycetes	226	226	91%	6.00E-71	50%	Phi.sp.-ADZ99360.1-Fun-Leo
177	XP_001935162.1	endoglucanase-5	Pyrenophora tritici-repentis P	Pyr.tri	Fungi	Leotiomycetes	225	225	86%	7.00E-71	50%	Pyr.tri-XP_001935162.1-Fun-Leo
178	EUC59305.1	glycoside hydrolase family 45 protein	Rhizoctonia solani AG-3 Rhs1A	Rhi.sol	Fungi	Agaricomycetes	225	225	90%	7.00E-71	53%	Rhi.sol-EUC59305.1-Fun-Aga
179	ORX52602.1	hypothetical protein BCR36DRAFT_58254	Piromyces finnis	Pir.fin	Fungi	Neocallimastigaceae	229	229	86%	1.00E-70	54%	Pir.fin-ORX52602.1-Fun-Neo
180	BAA98042.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	224	224	86%	1.00E-70	55%	Ret.spe-BAA98042.1-Euk-Par
181	CCA72362.1	probable endoglucanase	Serendipita indica DSM 11827	Ser.ind	Fungi	Agaricomycetes	226	226	99%	2.00E-70	52%	Ser.ind-CCA72362.1-Fun-Aga
182	ADU33249.1	endo-beta-1,4-glucanase	Sitophilus oryzae	Sit.ory	Hexapod	Curculionidae	224	224	98%	2.00E-70	49%	Sit.ory-ADU33249.1-Hex-Cur
183	ORX80640.1	endoglucanase 45A	Anaeromyces robustus	Ana.rob	Fungi	Neocallimastigaceae	228	228	86%	2.00E-70	54%	Ana.rob-ORX80640.1-Fun-Neo
184	XP_016642124.1	Glycosyl hydrolase family 45	Scedosporium apiospermum	Sce.api	Fungi	Sordariomycetes	224	224	87%	2.00E-70	52%	Sce.api-XP_016642124.1-Fun-Sor
185	KXH44598.1	glycosyl hydrolase family 45	Colletotrichum nymphaeae SA	Col.nym	Fungi	Sordariomycetes	224	224	90%	2.00E-70	51%	Col.nym-KXH44598.1-Fun-Sor
186	OUM56845.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	223	223	86%	3.00E-70	52%	Pir.sp.-OUM56845.1-Fun-Neo
187	ORX45359.1	hypothetical protein BCR36DRAFT_58567	Piromyces finnis	Pir.fin	Fungi	Neocallimastigaceae	229	229	87%	3.00E-70	55%	Pir.fin-ORX45359.1-Fun-Neo
188	AEE61695.1	unknown	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	224	224	96%	3.00E-70	48%	Den.pon-AEE61695.1-Hex-Cur
189	XP_001908770.1	hypothetical protein	Podospora anserina S mat+	Pod.ans	Fungi	Sordariomycetes	224	224	88%	3.00E-70	51%	Pod.ans-XP_001908770.1-Fun-Sor
190	XP_003652266.1	glycoside hydrolase family 45 protein	Thielavia terrestris NRRL 8126	Thi.ter	Fungi	Sordariomycetes	223	223	88%	3.00E-70	50%	Thi.ter-XP_003652266.1-Fun-Sor
191	XP_007916894.1	putative glycoside hydrolase family 45 pr	Phaeoacremonium minimum	Pha.min	Fungi	Sordariomycetes	224	224	98%	4.00E-70	48%	Pha.min-XP_007916894.1-Fun-Sor
192	CEG64031.1	Putative Endoglucanase	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	226	226	89%	4.00E-70	52%	Rhi.mic-CEG64031.1-Fun-Muc
193	OBZ88981.1	Endoglucanase-5	Choanephora cucurbitarum	Cho.cuc	Fungi	Mucorales	226	226	89%	5.00E-70	51%	Cho.cuc-OBZ88981.1-Fun-Muc
194	BAA98035.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	222	222	86%	5.00E-70	53%	Ret.spe-BAA98035.1-Euk-Par
195	OBT68825.1	hypothetical protein VE03_02186	Pseudogymnoascus sp. 23342-	Pse.sp.	Fungi	Leotiomycetes	225	225	87%	5.00E-70	52%	Pse.sp.-OBT68825.1-Fun-Leo
196	BAF57356.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	223	223	98%	5.00E-70	48%	unc.sym-BAF57356.1-Euk-Env
197	ORY36045.1	hypothetical protein LY90DRAFT_386889	Neocallimastix californiae	Neo.cal	Fungi	Neocallimastigaceae	227	227	90%	5.00E-70	54%	Neo.cal-ORY36045.1-Fun-Neo
198	XP_008089512.1	glycosyl hydrolase family 45	Colletotrichum graminicola M	Col.gra	Fungi	Sordariomycetes	223	223	87%	5.00E-70	53%	Col.gra-XP_008089512.1-Fun-Sor
199	BAA98047.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	222	222	86%	6.00E-70	52%	Ret.spe-BAA98047.1-Euk-Par
200	XP_019757792.1	PREDICTED: endoglucanase-like	Dendroctonus ponderosae	Den.pon	Hexapod	Curculionidae	223	223	98%	6.00E-70	47%	Den.pon-XP_019757792.1-Hex-Cur
201	KFH43153.1	putative endoglucanase type K-like prote	Acremonium chrysogenum AT	Acr.chr	Fungi	Hypocreomycetidae	223	223	86%	6.00E-70	53%	Acr.chr-KFH43153.1-Fun-Hyp
202	ORX45391.1	hypothetical protein BCR36DRAFT_33320	Piromyces finnis	Pir.fin	Fungi	Neocallimastigaceae	227	227	92%	6.00E-70	53%	Pir.fin-ORX45391.1-Fun-Neo
203	KFY56256.1	hypothetical protein V496_06744	Pseudogymnoascus sp. VKM F-	Pse.sp.	Fungi	Leotiomycetes	225	225	87%	6.00E-70	52%	Pse.sp.-KFY56256.1-Fun-Leo
204	XP_018156417.1	Endoglucanase-5	Colletotrichum higginsianum I	Col.hig	Fungi	Sordariomycetes	228	228	91%	6.00E-70	52%	Col.hig-XP_018156417.1-Fun-Sor
205	BAA98040.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	222	222	86%	7.00E-70	53%	Ret.spe-BAA98040.1-Euk-Par
206	XP_007686761.1	glycoside hydrolase family 45 protein	Bipolaris oryzae ATCC 44560	Bip.ory	Fungi	Dothideomycetes	222	222	86%	7.00E-70	49%	Bip.ory-XP_007686761.1-Fun-Dot
207	KXH61645.1	glycosyl hydrolase family 45	Colletotrichum salicis	Col.sal	Fungi	Sordariomycetes	222	222	86%	7.00E-70	52%	Col.sal-KXH61645.1-Fun-Sor
208	XP_003028771.1	glycoside hydrolase family 45 protein	Schizophyllum commune H4-8	Sch.com	Fungi	Agaricomycetes	222	222	87%	8.00E-70	54%	Sch.com-XP_003028771.1-Fun-Aga
209	EME46279.1	glycoside hydrolase family 45 protein	Dothistroma septosporum NZE	Dot.sep	Fungi	Leotiomycetes	222	222	86%	8.00E-70	51%	Dot.sep-EME46279.1-Fun-Leo
210	BAA98045.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	222	222	86%	8.00E-70	53%	Ret.spe-BAA98045.1-Euk-Par
211	KFY59844.1	hypothetical protein V497_04055	Pseudogymnoascus sp. VKM F-	Pse.sp.	Fungi	Leotiomycetes	232	232	87%	9.00E-70	51%	Pse.sp.-KFY59844.1-Fun-Leo
212	XP_018031233.1	glycoside hydrolase	Paraphaeosphaeria sporulosa	Par.spo	Fungi	Leotiomycetes	222	222	94%	9.00E-70	46%	Par.spo-XP_018031233.1-Fun-Leo
213	OTA68808.1	glycoside hydrolase family 45 protein	Hypoxylon sp. EC38	Hyp.sp.	Fungi	Sordariomycetes	221	221	88%	1.00E-69	48%	Hyp.sp.-OTA68808.1-Fun-Sor
214	CEI85881.1	Putative Endoglucanase	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	225	225	89%	1.00E-69	52%	Rhi.mic-CEI85881.1-Fun-Muc
215	XP_008077238.1	Barwin-like endoglucanase	Glarea lozoyensis ATCC 20868	Gla.loz	Fungi	Leotiomycetes	225	225	97%	1.00E-69	48%	Gla.loz-XP_008077238.1-Fun-Leo
216	ODM90131.1	Endoglucanase-5	Orchesella cincta	Orc.cin	Hexapod	Collembola	221	221	86%	1.00E-69	51%	Orc.cin-ODM90131.1-Hex-Col
217	ACM44323.1	cellulase BXC12	Bursaphelenchus xylophilus	Bur.xyl	Meta	Nematode	225	225	92%	1.00E-69	50%	Bur.xyl-ACM44323.1-Met-Nem

218	OUM61087.1	glycoside hydrolase family 45 protein	Piromyces sp. E2	Pir.sp.	Fungi	Neocallimastigaceae	221	221	86%	2.00E-69	56% Pir.sp.-OUM61087.1-Fun-Neo
219	XP_001220089.1	hypothetical protein CHGG_00868	Chaetomium globosum CBS 14	Cha.glo	Fungi	Sordariomycetes	221	221	97%	2.00E-69	47% Cha.glo-XP_001220089.1-Fun-Sor
220	XP_020050301.1	glycoside hydrolase family 45 protein	Ascoidea rubescens DSM 1968	Asc.rub	Fungi	Saccharomycetales	230	230	90%	2.00E-69	53% Asc.rub-XP_020050301.1-Fun-Sac
221	OCW43032.1	endoglucanase-5	Diaporthe helianthi	Dia.hel	Fungi	Sordariomycetes	224	224	97%	2.00E-69	49% Dia.hel-OCW43032.1-Fun-Sor
222	XP_007855174.1	glycoside hydrolase family 45 protein	Moniliophthora roreri MCA 29	Mon.ror	Fungi	Agaricomycetes	221	221	87%	2.00E-69	51% Mon.ror-XP_007855174.1-Fun-Aga
223	XP_001801548.1	hypothetical protein SNOG_11303	Parastagonospora nodorum SN	Par.nod	Fungi	Leotiomycetes	221	221	96%	2.00E-69	46% Par.nod-XP_001801548.1-Fun-Leo
224	XP_007840161.1	hypothetical protein PFICI_13389	Pestalotiopsis fici W106-1	Pes.fic	Fungi	Sordariomycetes	221	221	90%	3.00E-69	50% Pes.fic-XP_007840161.1-Fun-Sor
225	XP_006696369.1	cellulase-like protein	Chaetomium thermophilum va	Cha.the	Fungi	Sordariomycetes	221	221	86%	3.00E-69	52% Cha.the-XP_006696369.1-Fun-Sor
226	OTA97052.1	glycoside hydrolase family 45 protein	Hypoxylon sp. CO27-5	Hyp.sp.	Fungi	Sordariomycetes	221	221	88%	3.00E-69	48% Hyp.sp.-OTA97052.1-Fun-Sor
227	XP_016584462.1	hypothetical protein SPSK_00893	Sporothrix schenckii 1099-18	Spo.sch	Fungi	Sordariomycetes	221	221	96%	3.00E-69	49% Spo.sch-XP_016584462.1-Fun-Sor
228	OQV16017.1	Endoglucanase-5	Hypsibius dujardini	Hyp.duj	Meta	Tardigrada	226	226	88%	4.00E-69	53% Hyp.duj-OQV16017.1-Met-Tar
229	AKC91346.1	glycoside hydrolase 45	Adineta ricciae	Adi.ric	Meta	Lophotrochozoa	225	225	86%	4.00E-69	55% Adi.ric-AKC91346.1-Met-Lop
230	OIW24112.1	glycoside hydrolase	Coniochaeta ligniaria NRRL 30	Con.lig	Fungi	Sordariomycetes	221	221	92%	4.00E-69	48% Con.lig-OIW24112.1-Fun-Sor
231	BAF57325.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	220	220	86%	5.00E-69	52% unc.sym-BAF57325.1-Euk-Env
232	BAF57323.1	putative glycosyl hydrolase family45	uncultured symbiotic protist o	unc.sym	Eukar	Env-samples	219	219	86%	5.00E-69	52% unc.sym-BAF57323.1-Euk-Env
233	BAA98046.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	219	219	87%	6.00E-69	52% Ret.spe-BAA98046.1-Euk-Par
234	KZL80527.1	endoglucanase type k	Colletotrichum incanum	Col.inc	Fungi	Sordariomycetes	225	225	98%	8.00E-69	50% Col.inc-KZL80527.1-Fun-Sor
235	ADU33250.1	endo-beta-1,4-glucanase	Sitophilus oryzae	Sit.ory	Hexapod	Curculionidae	220	220	97%	8.00E-69	48% Sit.ory-ADU33250.1-Hex-Cur
236	BAD34543.1	beta-1,4-endoglucanase	Bursaphelenchus xylophilus	Bur.xyl	Meta	Nematode	219	219	99%	8.00E-69	46% Bur.xyl-BAD34543.1-Met-Nem
237	BAA98044.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	219	219	87%	9.00E-69	51% Ret.spe-BAA98044.1-Euk-Par
238	ORY57204.1	cellulase-like protein	Pseudomassariella vexata	Pse.vex	Fungi	Sordariomycetes	221	221	88%	9.00E-69	49% Pse.vex-ORY57204.1-Fun-Sor
239	KXL46237.1	glycoside hydrolase family 45 protein	Acidomyces richmondensis	Aci.ric	Fungi	Dothideomycetes	223	223	99%	1.00E-68	48% Aci.ric-KXL46237.1-Fun-Dot
240	CZR62396.1	related to endoglucanase	Phialocephala subalpina	Phi.sub	Fungi	Leotiomycetes	223	223	91%	1.00E-68	49% Phi.sub-CZR62396.1-Fun-Leo
241	ADU33286.1	endo-beta-1,4-glucanase	Chrysomela tremula	Chr.tre	Hexapod	Chrysomeloidea	219	219	84%	1.00E-68	52% Chr.tre-ADU33286.1-Hex-Chr
242	BAA98043.1	family 45 cellulase homologue	Reticulitermes speratus hindg	Ret.spe	Eukar	Parabasalia	219	219	87%	1.00E-68	51% Ret.spe-BAA98043.1-Euk-Par
243	KDQ31420.1	glycoside hydrolase family 45 protein	Pleurotus ostreatus PC15	Ple.ost	Fungi	Agaricomycetes	222	222	100%	1.00E-68	48% Ple.ost-KDQ31420.1-Fun-Aga
244	BAD34546.1	beta-1,4-endoglucanase	Bursaphelenchus xylophilus	Bur.xyl	Meta	Nematode	219	219	99%	2.00E-68	46% Bur.xyl-BAD34546.1-Met-Nem
245	XP_003714252.1	endoglucanase	Magnaporthe oryzae 70-15	Mag.ory	Fungi	Sordariomycetes	219	219	87%	2.00E-68	50% Mag.ory-XP_003714252.1-Fun-Sor
246	WP_084668122.1	hypothetical protein	Myxococcus stipitatus	Myx.sti	Bacteria	Deltaproteobacteria	219	219	86%	2.00E-68	51% Myx.sti-WP_084668122.1-Bac-Del
247	AOW70406.1	glycoside hydrolase family 45 protein	uncultured organism	unc.org	Unclass	Unclass	219	219	89%	2.00E-68	51% unc.org-AOW70406.1-Unc-Unc
248	CEG76406.1	Putative Endoglucanase	Rhizopus microsporus	Rhi.mic	Fungi	Mucorales	222	222	89%	2.00E-68	51% Rhi.mic-CEG76406.1-Fun-Muc
249	AGC43397.1	beta-1,4-endoglucanase	Myxococcus stipitatus DSM 14	Myx.sti	Bacteria	Deltaproteobacteria	219	219	86%	2.00E-68	51% Myx.sti-AGC43397.1-Bac-Del
250	ORE00783.1	endo-beta-D-1,4-glucanase	Rhizopus microsporus var. mic	Rhi.mic	Fungi	Mucorales	221	221	89%	2.00E-68	51% Rhi.mic-ORE00783.1-Fun-Muc
251	GEXX01020088.1	_	Achipteria coleoptrata	Ac.col	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Ac.col-GEXX01020088.1-Chel
252	GEXX01102535.1	_	Achipteria coleoptrata	Ac.col	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Ac.col-GEXX01102535.1-Chel
253	GEY01099922.1	_	Hermannia gibba	Her.gib	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Her.gib-GEY01099922.1-Chel
254	GEYB01090604.1	_	Hermannia gibba	Her.gib	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Her.gib-GEYB01090604.1-Chel
255	GEYB01106504.1	_	Hermannia gibba	Her.gib	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Her.gib-GEYB01106504.1-Chel
256	GEYB01110804.1	_	Hermannia gibba	Her.gib	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Her.gib-GEYB01110804.1-Chel
257	GEY01068052.1	_	Nothrus palustris	No.pal	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ No.pal-GEY01068052.1-Chel
258	GEYZ01031500.1	_	Platynothrus peltifer	Pla.ple	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Pla.pel-GEYZ01031500.1-Chel
259	GEYQ01020831.1	_	Steganacarus magnus	Ste.mag	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Ste.mag-GEYQ01020831.1-Chel
260	GEYQ01031992.1	_	Steganacarus magnus	Ste.mag	Arthropoda	Chelicerata/Oribatida	_	_	_	_	_ Ste.mag-GEYQ01031992.1-Chel
261	GAUE02010707.1	_	Anurida maritima	An.ma	Hexapoda	Collembola	_	_	_	_	_ An.ma-GAUE02010707.1-Hex-Col
262	GASX02031894.1	_	Folsomia candida	Fo.can	Hexapoda	Collembola	_	_	_	_	_ Fo.can-GASX02031894.1-Hex-Col
263	GAMM01001160.1	_	Orchesella cincta	Or.cin	Hexapoda	Collembola	_	_	_	_	_ Or.cin-GAMM01001160.1-Hex-Col
264	GATD02028261.1	_	Pogonognathellus sp.	Po.sp	Hexapoda	Collembola	_	_	_	_	_ Po.sp-GATD02028261.1-Hex-Col

Table S3. Details on the genome/transcriptome datasets which were used to curate GH45 sequences derived from Phytophaga beetles.

Species	Acronym	Superfamily	family	subfamily	Data type*	accession
Cylas brunneus	CBR	Curculionoidea	Brentidae	Brentinae	SRA	SRX1710181
Cylas formicarius	CFO	Curculionoidea	Brentidae	Brentinae	SRA	SRX1508049
Cylas puncticollis	CPU	Curculionoidea	Brentidae	Brentinae	SRA	SRX732288
Anthonomus grandis	AGR	Curculionoidea	Curculionidae	Curculioninae	SRA	SRX2888367
Listronotus oregonensis	LOR	Curculionoidea	Curculionidae	Cyclominae	SRA	SRX1674531; SRX1674530; SRX1674529
Cyrtotrachelus buqueti	CBU	Curculionoidea	Curculionidae	Dryophthorinae	SRA	SRX3262946
Rhynchophorus ferrugineus	RFE	Curculionoidea	Curculionidae	Dryophthorinae	TSA	GDKA00000000.1
Sitophilus oryzae	SOR	Curculionoidea	Curculionidae	Dryophthorinae	SRA	SRX017240
Sphenophorus levis	SLE	Curculionoidea	Curculionidae	Dryophthorinae	Sanger ESTs	JZ135722.1-JZ139168.1
Diaprepes abbreviatus	DAB	Curculionoidea	Curculionidae	Entiminae	Sanger ESTs	CN472512.1-CN488395.1; DN199437.1-DN201109.1
Pachyrhynchus infernalis	PIN	Curculionoidea	Curculionidae	Entiminae	SRA	DRX089461; DRX089463; DRX089465
Hylobius abietis	HAB	Curculionoidea	Curculionidae	Molytinae	SRA	SRX3423630
Larinus minutus	LMI	Curculionoidea	Curculionidae	Molytinae	TSA	GDLA00000000.1
Pissodes strobi	PST	Curculionoidea	Curculionidae	Molytinae	Sanger ESTs	GT285068.1-GT296156.1
Euwallacea fornicatus	EFO	Curculionoidea	Curculionidae	Scolytinae	SRA	SRX698946
Hypothenemus hampei	HHA	Curculionoidea	Curculionidae	Scolytinae	Genome	LBGY00000000.1
Ips pini	IPI	Curculionoidea	Curculionidae	Scolytinae	Sanger ESTs	CB407474.1-CB409136.1
Ips typographus	ITY	Curculionoidea	Curculionidae	Scolytinae	TSA	GACR00000000.1
Tomicus yunnanensis	TYU	Curculionoidea	Curculionidae	Scolytinae	SRA	SRX2518383
Dendroctonus ponderosae	DPO	Curculionoidea	Curculionidae	Scolytinae	Genome	APGK00000000.1
Anoplophora glabripennis	AGL	Chrysomeloidea	Cerambycidae	Lamiinae	Genome	AQHT00000000.2
Apriona japonica	AJA	Chrysomeloidea	Cerambycidae	Lamiinae	SRA	ERX387572-ERX387579
Monochamus alternatus	MAL	Chrysomeloidea	Cerambycidae	Lamiinae	SRA	SRX1302202; SRX1605798; SRX1605842; SRX1606425; SRX1606426
Anoplophora chinensis	ACH	Chrysomeloidea	Cerambycidae	Lamiinae	NR	AFN89565.1
Apriona germari	AGE	Chrysomeloidea	Cerambycidae	Lamiinae	NR	AAU44973.1, AAR22385.1
Batocera horsfieldi	BHO	Chrysomeloidea	Cerambycidae	Lamiinae	NR	AKH90729.1
Mesosa myops	MMY	Chrysomeloidea	Cerambycidae	Lamiinae	NR	AMA76413.1
Oncideres albomarginata chamela	OAL	Chrysomeloidea	Cerambycidae	Lamiinae	NR	ADI24132.1
Psacothea hilaris	PHI	Chrysomeloidea	Cerambycidae	Lamiinae	NR	ALE71518.1
Cassida rubiginosa	CRU	Chrysomeloidea	Chrysomelidae	Cassidinae	SRA	SRX3287590; SRX3287591; SRX3287593
Octodonta nipae	ONI	Chrysomeloidea	Chrysomelidae	Cassidinae	SRA	SRX396790
Chrysomela tremula	CTR	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRA	SRX017241

Chrysomela populi	CPO	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRA	SRX390590; SRX390602; SRX390603
Colaphellus bowringi	CBO	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRA	SRX317064
Gastrophysa viridula	GVI	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRA	SRX017237
Leptinotarsa decemlineata	LDE	Chrysomeloidea	Chrysomelidae	Chrysomelinae	SRA	SRX017239
Oreina cacaliae	OCA	Chrysomeloidea	Chrysomelidae	Chrysomelinae	TSA	GDPL00000000.1
Phaedon cochleariae	PCO	Chrysomeloidea	Chrysomelidae	Chrysomelinae	in-house transcriptome	
Clitea metallica	CME	Chrysomeloidea	Chrysomelidae	Galerucinae	SRA	SRX3921909
Diabrotica virgifera virgifera	DVI	Chrysomeloidea	Chrysomelidae	Galerucinae	TSA	GBSB00000000.1
Phyllotreta armoraciae	PAR	Chrysomeloidea	Chrysomelidae	Galerucinae	in-house transcriptome	
Podagricomela weisei	PWE	Chrysomeloidea	Chrysomelidae	Galerucinae	SRA	SRX3921907
Psylliodes chrysocephala	PCH	Chrysomeloidea	Chrysomelidae	Galerucinae	in-house transcriptome	

*SRA: Short read archive at NCBI; TSA: transcriptome shotgun assembly at NCBI; NR: non redundant protein database at NCBI.

Table S4. Complete list of primers used in this study

Primers used for full length amplifacation including Kozak Sequenz			
Species	Target Gene	5´Primer	3´Primer
C. tremula	GH45-1	ACC ATG GAG CTT TCA GTG GTT ATT TTG	GTA ATC ACA ACC GGT GAT TGC AA
	GH45-2	AGG ATG GAC TTT CTA GCT ATC GGT TC	ACT AAC TGG ATT ACA ACC GCT GA
	GH45-3	GAC ATG GAA TCA ATC ATT TTA GTT TCT G	TTC CAA ATC TCC ACA TCC GGT G
P. cochleariae	GH45-1	GAG ATG GAG GTT ATC GTG TTA CCC TTG	CAA TGC GCA ATT GGA GAT CGC CA
	GH45-3	AGG ATG GAA CTT TTA GTG ATT ATT TTG ACA	GTA ATC ACA TCC GGA AAT AGC AAC
	GH45-4	ACC ATG GAA TTA ATA ATT TTA GTT TCT GC	GTC CAA GTC TCC ACA TCC AGT A
	GH45-5	AGG ATG GAA TCA TTA ATT TTA GTT TCT GT	ATC CAA GTC TCC ACA TCC AGT A
	GH45-6	AGG ATG GAA CTA ATA GTT TTA GTT TCT GC	ATC CAA GTC TCC ACA TCC AGT GA
	GH45-7	AGG ATG GAC TTT CTA GCT ATT AGT TCA C	ACC AAC TGG ACT ACA ACC ACT AAT
	GH45-8	ACC ATG GAG CTT CTA GTG ATT ACA TTG G	ATA CTC GCA TCC TGT CAG CGC GA
L. decemlineata	GH45-1	GAA ATG GAA GGG TTT GTG GTA TTG TTG	ATA CTC GCA ACC TGT TAT TGC AAC
	GH45-2	GAC ATG GAA CTC TTG GTG ATA TCA TTG	ATC CCT ATC TCC ACA TTT TGA AAT ATC
	GH45-3	AGC ATG GGG CAC GTA ATT GTT TTG G	AAC ACG TAG CTC ACA GCC AG
	GH45-4	GAG ATG GTA ACT GCT TTG GCA TTC ATT	CTT CAA GTT ACA ATG AGA TCT GTC
	GH45-5	GAC ATG GAA ACT GCT TTA GCA CTC AT	CTT CAA GTT GCA ATG ACT TAT ATC
	GH45-6	GGA ATG GAG ATC GCA CTT CTG GCT T	CAA ATT ACA TCC GGT TAT CTC AAC
	GH45-7	AGC ATG GTC GTC AAT ATT ATT TCT TTC TT	CAT ATT ACA ACC AGT ATT GGC AAT C
	GH45-8	GAA ATG GAG CTT ATA GTA TTA TCT CTC G	ATC TCC ACA ACC AGT GAG AGC TT
	GH45-9	GAA ATG GAG CTT ATA GTA TTA TCT CTC G	ATC TCC ACA ACC AGT GAG AGC TT
	GH45-10	GAC ATG GAA CTC TTG GTG ATA TCA TTG	ATC CCT ATC TCC ACA TTT TGA AAT ATC
	GH45-11	ACC ATG GAA ACC ATA CCA ACG ATT CTG	ATT CAA AGG GAA ACA GCC AGT C
D. virgifera	GH45-1	GAG ATG GAG ATC GCT ATC CTT GTT TC	AAG ACT GCA ACC AGA TCT GGC A
	GH45-2	AGG ATG GAG GCA CAA CTT ATT TCA TTG	AAG TGG AAG ACT ACA TCC ACT AAT
	GH45-3	ACG ATG GTT CCT CTA CCT ATA CTT TTA G	CAA TAC ACA ACC ACT TCT TTC AAC A
	GH45-4	ACG ATG GTT TTT ATT ATA TTC TCT TTG CTA	TTC GGG AAT TCT ACA ACC ACT AAT

	GH45-5	GCG ATG GAA TAT ACA ATT ACT TCT CTG TT	AGG ATG AGA AAT GGG ATC ACA AC
	GH45-6	GCT ATG GAA TTC GCA GTG GCA TTT TTA	AGG ATG AGA TGC GGG ATA ACA G
	GH45-7	GAC ATG GAT ACC GGT ATA GTG AAT ATT TT	AAG ATC TCC ACA GCC TGT GAT G
	GH45-8	GCCACCATGGAATATTTAGTTGTTATTGCCT	ATCCAAGTCACCACAAGCAGA
	GH45-9	GCCACCATGGTGTCTTTGAAAATAGCATTTG	CAAATCTCCACATTTACTGATTC
	GH45-10	GCCACCATGGAAACGTTTACAGTTTTTGCTT	ATCTCTATCGCCACAGTTCGT
	GH45-11	GCCACCATGGTGTCTTTGAAAATAGCATTTG	CAAATCTCCACATTTACTGATTC
S. orizae	GH45-1	GAC ATG GAA GTT TTA TGC GTG ATT TTA G	TAC AAC GCA TCC GGT GAT ACT AG
	GH45-2	GTG ATG GAA CTT TTA TGC TTA TTT TTA GC	AGC ACA TCC GGT AAT ACT GGT C
	GH45-3	GTC ATG GAA GTG TTA ATT GTT TTG ACT A	ATC CGT TTT ACA GCC TGA GAT ATC
	GH45-4	GCT ATG GAA GTG TTA ATT GTT TTG ACT A	ATC CAT TTT ACA GCC TGA GAT ATC
	GH45-5	AGG ATG GAA GTT CTA ATT GTT TTA ACC TT	CGA ATT CCA AAC AGG GTC CAT T

Primers used for qPCR Amplification			
Species	Target Gene	5´Primer	3´Primer
P. cochleariae	GH45-1	TCC TTG TGC AGA TCA CCA AC	GTG ACA GCC CTG TGT GAA GA
	GH45-3	AGA CGG ATC TGG TGA AGC AT	CTT CAC CTC CCG TGA AAG AG
	GH45-4	CAA TCA TCC CAA TCG AAG GT	TGG TGT GGA TGA TTT TGT CC
	GH45-5	GGA GGA GGA GTG TAA CCA G	ACT AGG ACA TCG CAC CTG C
	GH45-6	ATT AAC TCC ACG CTG GCC TA	CTA GCT GCC CCT TGA ATG AC
	GH45-7	TGC TCG TAC AGG TGA CCA AC	CAG CCC TGC GAG TAG TAA C
	GH45-8	CAC TGG TGG GGA GGA TTA C	GTG TTC ACC ACT TGG ACG AG

Primers used for RACE-PCR			
Species	Target Gene	5´Primer	3´Primer
D. virgifera	GH45-11	CTG GGA CTG CTG CGC TCC ATC A	

4. General Discussion

The survival of plant-feeding organisms is dependent on a functional and diverse set of plant cell wall degrading enzymes (PCWDEs) either being provided endogenously or supplied by mutualistic symbionts. Most of our knowledge about PCWDEs is based on how they function in microbes. There they have been shown to be essential for breaking down the plant cell wall to provide themselves (and potential host species during symbiosis) with carbohydrates, a source of metabolic energy (Nguyen et al. 2018; Slaytor 1992). They are also used to penetrate the cell wall during plant infection of pathogens (Goyal et al. 1991; Joko et al. 2014; Kubicek et al. 2014) and to provide plant-feeding animals with tools to efficiently degrade the cell wall. Yet, little is known about the function, physiological importance or evolution of endogenous PCWDEs found in animals, especially in leaf beetles (Chrysomelidae) and weevils (Curculionidae). In this thesis, I investigate a set of PCWDEs encoded by several phytophagous beetles including six chrysomelid and one curculionid species. My focus is the glycoside hydrolase family 5 subfamily 10 (GH5_10s) and family 45 (GH45s), particularly with respect to their evolution, enzymatic properties and physiological importance.

I discovered that GH5_10 proteins, which to date are only known to have been encoded by two chrysomelid species within the Phytophaga, have kept their ancestral mannanolytic activity. Strikingly, one of the four copies of GH5_10s in *C. maculatus* has lost its mannanase activity but has evolved to degrade xylan, whereas the GH5_10 mannanase has evolved to also degrade cellulose (**Manuscript I**). Likewise, I have learned from my investigation of the GH45s in *G. viridula* that one of them has lost its original cellulolytic function and evolved to become an endo-β-1,4-xyloglucanase, while the other kept its ancestral endo-β-1,4-glucanase activity (**Manuscript II**). The evolution of novel substrate specificities in these proteins likely compensates for the absence of other glycoside hydrolase families typically covering the respective enzymatic activities. More precisely *C. maculatus* GH5_10 has evolved cellulolytic as well as xylanolytic activity due to the absence of GH families with the corresponding activity pattern (such as GH10/GH11 xylanases or GH45/GH9

cellulases). Likewise, GH45 xyloglucanase activity in *G. viridula* may has evolved because no other GH family is encoded by the genome of this beetle being able to degrade xyloglucan. Surprisingly, I was unable to detect any changes in mortality or weight gain after GH5_10 or GH45 knockdown in *G. viridula* indicating a compensation by other digestive enzymes encoded by the insects (**Manuscripts I and II**). Furthermore, I present data on the complexity of GH45 and GH5_10 evolution within animals and more specifically within beetles. Although I have been able to show that the GH5_10 family has been acquired at least twice within insects, the ancestral donor organism has remained elusive (**Manuscript I**). Likewise, I have discovered that the GH45 family was not inherited ancestrally but has likely been acquired at least three times independently in arthropods. The sister clade to beetle-derived GH45s was composed of fungal sequences, indicating a likely horizontal gene transfer (HGT) event from fungi to beetle (**Manuscript III**).

4.1 PCWDEs in Animals: Impact on Science, Survival and Society

The following paragraphs will focus on the effect of Coleoptera-derived PCWDEs on science, beetle survival and the benefit of PCWDEs for society.

4.1.1 Scientific impact – PCWDE evolution

After the first endogenous cellulases were found in animals, the scientific world was confronted with its dogmatic concept of a cellulolytic system only present in microbes. Yet, the initial surprise was quickly overcome as this “brave new world” of cellulolytic systems in animals initiated new research areas. The main question that arose was, how have cellulases evolved in animals? To find an answer, researchers started to include phylogenetic analyses and other tools to investigate the evolutionary origin of cellulases as well as other PCWDEs. Two major hypotheses began to gain shape; either an ancestral evolution of genes encoding PCWDEs or the genes were acquired from a foreign source by HGT. Surprisingly, both hypotheses seemed to be true. For instance, an ancestral origin has been proposed for cellulases of the GH9 family, which were likely present in the last common ancestor (LCA) of all Eukaryota and Bacteria (Davison and Blaxter 2005), whereas cellulases of GH5_2 were likely acquired from a bacterial source during the evolution of longhorned beetles

(Cerambycidae) and nematodes (Danchin et al. 2010). Over time, independent studies have suggested that the acquisition of PCWDEs by HGT in animals was more common than expected. For example pectinases of family GH28 have been shown to be initially acquired by Phytophaga beetles from a fungal source (Kirsch et al. 2014; Kirsch et al. 2016) followed by two additional independent HGTs from yet another fungal source as well as from bacteria; cellulases of family GH45 were acquired from a fungal source by nematodes (Kikuchi et al. 2004), molluscs (Sakamoto and Toyohara 2009) and the bdelloid rotifer *Adineta ricciae* (Szydlowski et al. 2015). In concert with a foreign origin of GH45s in nematodes and molluscs, my own research on the respective gene family suggests that it has been acquired by arthropods at least three times (**Manuscript III**). Based on this growing body of evidence, the evolution of glycoside hydrolases in insects indicates that their origin may be less conservative than anticipated. Thus, in the near future other cases of horizontally transferred PCWDEs encoded by animals are likely to be found.

In the meantime, several other intriguing questions remain unaddressed by the scientific community, e.g. why is it that beetle-derived GH45s and GH5_10s occur exclusively as “single“ enzymes? In contrast, their microbial counterparts are expressed as multi-enzyme complex proteins (MEC) in fungi or as cellulosomes in bacteria (Kuhad et al. 2011b; Leschine 1995). To date plant-cell-wall-degrading MEC in animals are restricted to some nematodes (GH5_2) (Danchin et al. 2010). The lack of MEC in insects has been further emphasized by my own research on the PCWDEs of phytophagous beetles (**Manuscripts I, II, and III**). Yet, the efficiency with which MEC or cellulosomes degrade cellulose is well documented and thus begs the question, why are they not present in beetles? Carbohydrate-binding modules (CBMs; part of MEC and cellulosomes) are believed to facilitate substrate cleavage, especially on crystalline cellulose, by increasing enzyme concentration in the proximity of the substrate (Bolam et al. 1998) and by disrupting the crystalline structure of cellulose (Boraston et al. 2004; Guillen et al. 2010). Thus, I hypothesize that the lack of CBMs (and ultimately MEC) in beetles may be due to low amounts of crystalline cellulose in the food source of target beetles, a source made up of either leaves or seeds. Based on the ability of both food sources to grow or initiate growth,

primary cell wall content is likely to be high. Although the ratio of amorphous to crystalline cellulose in the primary cell wall is still an issue among plant biologists (Ruel et al. 2012), it is suggested that the primary cell wall is comprised of less crystalline cellulose than secondary cell walls (Cosgrove 2014; Montanari et al. 2005). Thus, crystalline cellulose may be underrepresented in either leaves or seeds compared to its amorphous counterpart. Higher amorphous cellulose content in leaves or seeds may be further supported by the fact that most characterized beetle-derived cellulases are endo-β-1,4-glucanases, which are able to degrade only amorphous cellulose (**Manuscripts I, II, III**). Thus, the ratio of crystalline to amorphous cellulose is likely to favor the latter in leaves as well as seeds and may explain why cellulosomes (or MECs) are redundant in beetles. Notably, some plant pathogenic fungi and bacteria infect the same tissues (leaves) eaten by beetles, yet they still express cellulosomes and MEC. Because microbes cannot grind and crush leaf material, they are depend on a highly efficient set of PCWDEs, including cellulases. Thus, cellulosomes and MECs are necessary to quickly and thoroughly degrade cellulose (independent of its amorphous or crystalline state) in order to effectively infect plant tissue.

Notably, the proposed low amounts of crystalline cellulose do not appear to characterize the food source of beetles of the Cerambycidae (longhorned beetles), which are known to feed on woody plant tissue with high crystalline cellulose content. Yet, most characterized cellulases in cerambycid beetles are endo-active, single enzymes (Calderon-Cortes et al. 2010; Lee et al. 2005; McKenna et al. 2016; Mei et al. 2015; Pauchet et al. 2014a). Here, crystalline cellulose may be addressed by the fact that beetles masticate their food source before digestion. Chewing enlarges the substrate's surface and crystalline cellulose components may be hydrated, allowing enzymes to access and digest former crystalline regions. Additionally, the acidic milieu in the gut of phytophagous beetles (Sinha 1959b) may be able to dissolve parts of the crystalline cellulose structure into its amorphous state; a strategy that is also used *in vitro* by scientists to produce artificial amorphous cellulose from its crystalline counterpart (**Manuscripts II,III**). If true, the endogenous cellulolytic system of beetles of the Cerambycidae does not need to address crystalline cellulose by enzymatic

means and likely presents an alternative option of coping with crystalline cellulose without the support of an MEC. Finally, microbes digesting plant material are often anaerobic (Leschine 1995). The ratio of adenosine triphosphate (ATP) generated from glucose compared between anaerobic and aerobic microbes favors the latter (2 ATP in anaerobes compared to 38 ATP in aerobes). Therefore, a quick and efficient digestion of crystalline as well as amorphous cellulosic material in anaerobic bacteria (using cellulosomes) is necessary in order to release the amount of glucose needed for conversion to ATP (Goyal et al. 1991). In beetles the degradation of amorphous cellulose down to glucose, in combination with the released sugar-rich cell content (**Manuscript II**), most likely provides enough ATP given the insects' aerobic metabolism.

Collectively, the absence of those MECs and cellulosomes in beetles may be connected to the putatively lower crystallinity of cellulose in the plant cell wall of their host plants and/or to the insects' ability to masticate their food source before digestion. By degrading amorphous cellulosic regions, the resulting glucose moieties together with the nutrient-rich cell content are most likely sufficient to sustain beetle energy metabolism. However, what can be said is that the digestion of crystalline cellulose probably plays a less important role in beetles than the digestion of its amorphous counterpart. It remains interesting whether future research will unravel PCWDEs encoded by beetle which are incorporated into MECs and whether these enzymes are able to degrade crystalline cellulose.

4.1.2 Animal impact – PCWDE and survival

As has been discussed in the research presented here, enzymatic activity and ancestral origin suggest that GH45s were likely an essential prerequisite for the evolution of Phytophaga beetles' ability to feed on plants (**Manuscripts II, III**). However, an important question remains unaddressed. Did these beetles feed on plant material before the potential acquisition of PCWDEs? If yes, did these beetles feed on plant material with the support of mutualistic PCWDE-secreting symbionts, or, if no, what was their original food source? Interestingly, symbiotic microbes have been proposed as a crucial prerequisite for insects and their evolutionary transition to

herbivory (Hansen and Moran 2014; Russell et al. 2009). This proposed symbiosis would ensure a close proximity of the receiver species (the beetle) of genes encoding PCWDEs and a potential donor species (a microbe), a physiological necessity if our claim of an HGT of GH45s is true. Therefore, a potential symbiotic relationship between an ancestral beetle species and a PCWDE-encoding microbe appears plausible. Nonetheless, if this hypothesis holds true, one would expect some PCWDE-secreting symbionts to be harbored by phytophagous beetles. Surprisingly, reports of the proposed symbionts are scarce and restricted to a single beetle, *Cassida rubiginosa* (Chrysomelidae: Cassidinae), which shelters symbionts encoding a GH28 pectinase (Salem et al. 2017). Additionally, symbiotic microbes are known from other insects including termites (Breznak and Brune 1994) and wood-feeding cockroaches (Dietrich et al. 2014); however to the best of our knowledge no HGT of PCWDEs has been observed in either insect clade. Interestingly, these two clades of insects encode GH9 putative cellulases, which are assumed to be of ancestral origin (Davison and Blaxter 2005). Here, symbiont-derived PCWDEs act synergistically with endogenous GH9s to degrade plant tissue (Brune 2014; Ni and Tokuda 2013).

Another hypothesis proposes that the ancestral feeding habit within Coleoptera comprised saprophagy and fungivory (Schigel 2012), suggesting that the common ancestor of Phytophaga beetles ingested fungi and/or decaying wood matter. Saprotrophic fungi are omnipresent decomposers of decaying wood litter and encode a set of PCWDEs (Payne et al. 2015). Within Coleoptera, fungivory can still be found in ambrosia beetles (Cuculionidae: Scolytinae and Platypodinae) (Mueller et al. 2005) and in closely related beetles of the superfamilies Cucujoidea (Jacob 1996; Silva and Lapenta 2011) and in the more distantly related Tenebrionoidea (Majka 2007). By ingesting decaying plant tissue, the ancestor of the Phytophaga may have come in close contact with saprotrophic microbes; these could be the subsequent basis for a potential GH45-HGT event between beetle and fungus. In summary, symbiont-supported herbivory of early beetles appears to be equally as likely as the saprophagy/fungivory that occurred prior to the acquisition of PCWDEs in the ancestor of Phytophaga beetles and may be solved in the future when other symbiont-harboring beetles are discovered.

Consequently, the question arises regarding the benefit of encoding an endogenous set of PCWDEs instead of these being provided by mutualistic symbionts. Presumably, the major benefit of encoding an endogenous set of PCWDEs lies in the expansion of those gene families that leads to an increased gene copy number, which in turn leads to an increased mutational system robustness and ultimately allows for the acquisition of novel enzymatic functions by concomitantly maintaining ancestral functions (Wagner 2008). Gene expansion events of several GH families encoding PCWDEs have been shown by our group for GH28s (Kirsch et al. 2014), GH5_2 (Pauchet et al. 2014a), GH5_10 (**Manuscript I**) and GH45s (**Manuscript III**). These gene expansions were always coupled with enzymatic sub-functionalization that allow, e.g., GH5_2 to degrade xylan and xyloglucan in addition to cellulose (McKenna et al. 2016; Pauchet et al. 2014a); GH28s to degrade polygalacturonic oligomers in addition to homogalacturonan; and GH45s to degrade xyloglucan and glucomannan in addition to cellulose (**Manuscripts II and III**). In contrast, the expansion of the bacterial gene family is mainly driven by HGT, resulting in a slower sub- and/or neo-functionalization rate compared to the rate in animal counterparts (Treangen and Rocha 2011). Thus, the acquisition of an endogenous set of PCWDEs in the LCA of Phytophaga beetles, followed by gene duplication and functional diversification, likely led to a digestive system which allowed beetles to adapt faster to novel food sources than their (putative) symbiont-containing ancestral beetle counterpart. Notably, our research on GH45s revealed fungi as putative donor source. Although gene expansion mechanisms in fungi is more similar to other eukaryotes than to bacteria (Kelkar and Ochman 2012), the number of individual PCWDE-encoding GH families is low (Payne et al. 2015). Therefore, beetles with a potentially acquired set of PCWDEs may have been able to adapt better to novel food sources than beetles containing symbiotic fungi with a less diversified set of PCWDEs. Finally, beetles that encode an endogenous set of PCWDEs acquire full regulatory control of the respective genes, which allows them to adjust gene expression and consequently protein amount to their specific needs rather than to that of the symbionts. Also, insects encoding their own set of PCWDEs may be less susceptible to the abiotic and biotic stresses which can negatively influence the microbial gut community.

4.1.3 Impact on Society - Industrial Applications

Several industries have found intriguing ways to involve PCWDEs (especially cellulases) that can be beneficial for society. To date they are being used in, for example, cotton processing, paper recycling, juice extraction and biofuel production, and as detergent enzymes and as animal feed additives (Wilson 2009). The following paragraph will focus on the potential application of GH45 and GH5_10 enzymes in several of those industries.

As the world is becoming more and more aware of dwindling fossil fuel resources and the increasing greenhouse effect, the call for alternative fuels is becoming louder by the year. Since the energy crisis in the 1970s, research has focused on the use of (ligno)cellulosic biomass as a renewable feedstock that occurs as waste products from agriculture or forestry (Li et al. 2014; Obeng et al. 2017). The industrial sector quickly realized its potential use for second-generation biofuel production as long as the entrapped cellulose could be efficiently converted into glucose (Sims et al. 2010). Consequently, PCWDEs have gained major public attention because they are able to efficiently degrade plant biomass polysaccharides into sugar monomers and these in turn can be fermented by microbes into fuel-supplemented ethanol or (more rarely) butanol. Therefore, a highly effective set of PCWDEs is needed to degrade cellulosic plant biomass including exo- and endo-cellulases as well as hemicellulases (Gusakov 2013; Li et al. 2014). The search by the biofuel industry for cellulolytic as well as hemicellulolytic enzymes may represent an opportunity for beetle-derived GH45 and GH5_10 enzymes to be used in this field. Our research has shown that these proteins efficiently degrade cellulosic as well as hemicellulosic plant cell wall compounds (**Manuscripts I, II, III**). However, currently cellulases are biosynthesized on a commercial scale almost exclusively from aerobic fungi such as *Trichoderma reesei* or *Humicola insolens*. Using microbes as bioreactors for cellulase biosynthesis offers the ability to bioengineer strains which express high yields of cellulases and provide specific activity towards crystalline cellulose (Obeng et al. 2017; Schulein 1998). In fact, the enzyme cocktail produced by *T. reesei* is dominated by exo-cellulases (Garvey et al. 2013). Thus, biomaterial with lower cellulose crystallinity may be degraded less efficiently without an adequate set of exogenously supplemented

endoglucanases. The *T. reesei* cellulolytic system may be compensated for by beetles-derived GH45 endoglucanases which can improve the degradation of various plant cell wall compositions. In a similar approach, externally provided β-glucosidases are supplemented from yet another organism based on *T. reesei´s* few endogenous β-glucosidases (Garvey et al. 2013; Obeng et al. 2017).

Other molecular approaches of involving beetle GH45/GH5_10 enzymes in plant biomass processing may be investigated. For example, it has been shown that yeasts are able to express three recombinant cellulase proteins simultaneously (Matano et al. 2013; Parisutham et al. 2014). Therefore, expressing a beetle-derived endo-cellulase together with an exo-cellulase and a β-glucosidase would provide a complete set with which to promote cellulolytic synergism. Additionally, GH45s have the potential to be included in artificial cellulosomes. These so-called designer cellulosomes incorporate PCWDEs from different species and apply them synergistically (Stern et al. 2015; Vazana et al. 2012). So far all the above-mentioned strategies have focused only on cellulolytic enzymes. Because hemicelluloses are tightly interwoven with cellulose, hemicellulolytic enzymes need to be involved in order to effectively degrade plant biomass (Garvey et al. 2013). Therefore, approaches similar to those described for GH45 cellulases may be conducted with beetle-derived GH45 mannanases/xyloglucanases as well as with GH5_10 enzymes.

Furthermore, major applications of exogenously added endo-cellulases and xylanases are to be found in wine and brewing industries; these increase wine quality and stability; improve coloring and filtration; and clarify must (Kuhad et al. 2011a). Likewise, in beer brewing, endoglucanases are used to improve yield and quality by decreasing wort viscosity and the degree of polymerization. In food industries, supplementing feed grains with cellulases and xylanases has been shown to improve the performance of cows (Dhiman et al. 2002). In addition, it has been shown that piglets and chicken that have ingested PCWDEs as food supplements have increased growth rates compared to the control treatment (Bhat 2000; Karmakar and Ray 2011; Saleh et al. 2005; Shrivastava et al. 2011). Beetle-derived cellulolytic and hemicellulolytic enzymes may also be applied in pulp and paper industries where,

among other features, they decrease fiber coarseness and pulp viscosity. During the refining and grinding of woody tissue, the use of cellulolytic enzymes has resulted in an energy saving of up to 40 % in contrast to purely mechanical pulping (Kuhad et al. 2011a; Pere et al. 2001). Endoglucanases and xylanases have also been shown to increase the efficiency of the de-inking process of paper making, which lowers environmental pollution (Chander Kuhad et al. 2010). Several other industrial processes make use of cellulolytic and hemicellulolytic enzymes by using them e.g. during oil and carotenoid extraction to improve yields, to clean cotton textiles and to improve crops growth by increasing soil fertilization (Kuhad et al. 2011a).

4.2 Conclusion

The discovery of PCWDEs encoded by phytophagous beetles is an extraordinary example of how animals can feed on plants without much help from mutualistic symbionts. In this thesis, I have provided evidence that the GH45 family was most likely not an ancestral gene family but, rather, evolved independently at least three times in arthropods. My data show that beetle GH45s are closely related to fungal GH45s implying that a gene transfer has occurred, likely from a fungal donor to the last common ancestor of the Phytophaga clade of beetles. This finding was further supported by similar discoveries in nematodes and mollusks which likewise originated from a fungal donor (Kikuchi et al. 2004; Sakamoto and Toyohara 2009). Whether those fungi were symbionts in the LCA of Phytophaga beetles is unclear and may be addressed when more beetle and fungal species are investigated in the future.

In accordance with my functional data, I hypothesize that these genes were (and are) essential for these beetles to feed on plants. The enzymatic diversification of GH45s and GH5_10s increased their synergism and allowed for additional components of the plant cell wall to be degraded. This synergism is further enhanced by the presence of GH28 pectinases in the beetle species investigated. Ultimately, all GH proteins encoded by the beetles allow for the plant cell wall to be degraded sufficiently in order to retrieve energy-rich sugar monomers as well as to gain access to the nutrient rich cell content.

Interestingly, and in contrast to the above-stated hypothesis, the silencing of individual GH45s and GH5_10s did not result in any significant change in phenotype (i.e. mortality and weight gain). This strongly suggests the presence of a robust system of PCWDEs in the beetle gut which may be connected to the synergy with other expressed digestive enzymes. Based on that hypothesis, I propose that the simultaneous knockdown of several PCWDEs with different substrate specificities would likely result in an alteration of the beetle performance.

My results show that GH45s have expanded independently and species-wise after their initial acquisition in the LCA of the Phytophaga clade of beetles. The observed gene expansion was very likely the foundation of the rise of novel enzymatic functions in beetle-derived GH45 and GH5_10 proteins. In turn, gene expansion and the evolution of novel enzymatic functions have likely facilitated the adaptation to novel host plants and in a broader perspective may have influenced the radiation of Phytophaga beetles at an early stage of their evolution.

5. Summary

The first line of defense against biotic and abiotic stresses in plants consists of a diverse set of sugar-based compounds forming the plant cell wall. The major component of the plant cell wall is cellulose, a polysaccharide consisting of β-1,4-linked glucose moieties. Any organism being able to degrade cellulose would benefit from a huge source of energy as well as gaining access to nutrient-rich cell contents. Cellulose-degrading enzymes (cellulases) are well described in a wide range of microbes but were thought to be absent in animals. Recently, it became clear that some animals encode endogenous plant cell wall degrading enzymes (PCWDEs) belonging to several glycoside hydrolase families (GH), including putative cellulases of family 45 (GH45) and putative mannanases of family 5 subfamily 10 (GH5_10). In Arthropoda, GH45s and GH5_10s are most prominently encoded by insects including the Phytophaga clade of beetles (leaf beetles, longhorned beetles, bark beetles and weevils). Nonetheless, the distribution of both GH families in insects is erratic and it is assumed that they are not of ancestral origin but were acquired separately, likely through horizontal gene transfer (HGT) events from microbe to animal. Despite the intricate evolution of GH45s and GH5_10s and an emerging role of PCWDEs in biofuel industries both GH families in Phytophaga beetles remain largely unexplored. Therefore, the major aim of this thesis was to investigate beetle-derived members of GH5_10 and GH45 with focus on their enzymatic activity, physiological importance and evolutionary history.

(i) The functional characterization of GH5_10 encoded by *Gastrophysa viridula* (encodes one gene copy) and *Callosobruchus maculatus* (encodes four gene copies) revealed that they have kept their ancestral function allowing them to degrade gluco- as well as galactomannan. Surprisingly, one GH5_10 copy of *C. maculatus* has lost its mannanolytic activity but has evolved the ability to degrade xylan instead. Moreover, the *C. maculatus* GH5_10 endo-β-1,4-mannanase was able to additionally degrade cellulose which has likely evolved due to the absence of any other gene family encoding a cellulolytic enzyme in

this beetle. Investigation of the physiological function of the GH5_10 mannanase from *G. viridula* using gene silencing through RNA interference (RNAi) resulted in a successful gene knockdown as well as a drastic reduction of the amount of the corresponding protein. However, changes in mortality or weight gain were not detected suggesting a robust system of PCWDEs in the beetle gut which may be connected to the synergy with other expressed digestive enzymes. Intron/exon structure analysis of beetle-derived GH5_10 genes suggests that the last common ancestor of these two beetles already possessed at least one GH5_10 gene. Moreover, the evolutionary history of GH5_10 proteins revealed that this gene family was limited to some Bacteria, Mollusca and Arthropoda other than beetles. Interestingly, insect-derived GH5_10s were not monophyletic but clustered within two distinct arthropod clades indicating that GH5_10 genes were acquired at least twice during insect evolution. Although the donor organism is still elusive, this data suggests that the GH5_10 family was acquired by insects through horizontal gene transfer.

(ii) *Gastrophysa viridula* was used as a model to investigate the enzymatic characteristics of two GH45s as well as their biological function. Functional characterization experiments revealed that one GH45 possessed endo-β-1,4-glucanase activity which corresponded with activity patterns observed in microbes. Intriguingly, the second GH45 encoded by *G. viridula* has lost its cellulolytic function but has evolved to degrade xyloglucan. After successful gene silencing of both GH45s, no striking changes regarding mortality or weight gain were detected. Based on a zymography experiment, several other cellulolytic enzymes were detected in the beetle which most likely compensate for the silenced GH45 cellulase, thus causing the lack of phenotype. In contrast, the same experiment did not reveal any other xyloglucanase in the beetle gut suggesting a likewise robust system of PCWDEs as observed for GH5_10. These experiments have demonstrated that both GH45s in *G. viridula* are active PCWDEs. One of them is able to degrade amorphous cellulose where the other has evolved to degrade xyloglucan. Yet, and similar to GH5_10, their biological role for the survival of beetles remains elusive.

(iii) As GH45 were barely investigated in beetles and were functionally characterized almost exclusively in Cerambycidae, GH45 functional characterization was spread to four additional species of Chrysomelidae as well as one species of Curculionidae. In correlation with the results from *G. viridula*, at least one GH45 endo-β-1,4-xyloglucanase as well as two endo-β-1,4-glucanases in each target beetle species was detected. Moreover, several GH45 endo-β-1,4-glucanases also accepted glucomannan as substrate. Phylogenetic analysis of beetle-derived GH45s revealed that the LCA of Phytophaga beetles possessed at least one GH45 cellulase. Furthermore, the substrate shift from cellulose to xyloglucan likely occurred through a substitution of an aspartate by a glutamate in catalytically important sites. However, the substitution happened at two different amino acid positions between Chrysomelidae and Curculionidae, suggesting that xyloglucanase activity evolved independently in both beetle families. Finally, analysis of the evolutionary history of animal-derived GH45s uncovered at least three distinct origins of GH45 within arthropods. The closest related clade to beetle GH45s contained fungal-derived sequences, suggesting a HGT event from fungi to beetle. In summary, within Arthropoda GH45s have evolved at least three times and were likely acquired by beetles through horizontal gene transfer from a fungal source. After several gene duplication events, some GH45s acquired novel enzymatic functions through amino acid substitutions in catalytically important amino acid residues.

In conclusion, this thesis has greatly contributed to our understanding of PCWDEs encoded by Phytophaga beetles. In particular, our knowledge on GH45 and GH5_10 members encoded by Chrysomelidae and Curculionidae has greatly increased, demonstrating that not a vertical but a horizontal gene transfer was likely responsible for GH45 (and possibly for GH5_10) inheritance in these beetles. The following species-specific, independent gene duplications allowed for functional diversification and likely adaptation to their food source. These results provide fundamental insights into the evolution of PCWDEs and the molecular mechanisms of acquiring novel enzymatic functions. Furthermore, based on a variety of industrial applications of PCWDEs, beetle-derived GH45 and GH5_10 enzymes may contribute greatly to

Society by being introduced into several industrial applications (e.g. biofuel production) and ultimately reducing a progressing greenhouse effect.

6. Zusammenfassung

Die erste Abwehr von Pflanzen gegen biotische und abiotische Stressfaktoren bildet die pflanzliche Zellwand, einem Polymer zusammengesetzt aus diversen Zuckerderivaten. Der Hauptbestandteil der Pflanzenzellwand besteht aus Zellulose - einem Polysaccharid aufgebaut aus β-1,4-verbundenen Glucoseeinheiten. Ein Organismus, der die Fähigkeit besitzt Zellulose abzubauen, würde sich damit nicht nur eine große Energiequelle erschließen, sondern sich auch Zugang zu einem nähstoffreichen Zellinhalt verschaffen. Zelluloseverdauende Enzyme (Zellulasen) waren lange Zeit dafür bekannt Bestandteil einiger Mikroorganismen zu sein, galten aber gleichzeitig als nicht von Tieren selbst zu kodierend. In den vergangenen zwei Jahrzehnten wurde jedoch bekannt, dass Pflanzenzellwand-verdauende Enzyme (PZVE) auch von einigen Tierarten endogen kodiert werden. Zu diesen PZVE gehören u.a. Vertreter der Glycosidhydrolasefamilie (GH) und schließen auch putative Zellulasen der Familie 45 (GH45) und potentielle Mannanasen der Familie 5 Unterfamilie 10 (GH5_10) mit ein. In Arthropoden sind GH45 und GH5_10 hauptsächlich innerhalb der Insekten vertreten, zu denen auch Käfer der Klade Phytophaga (Blatt-, Bock-, Borken- und Rüsselkäfer) zählen. Dennoch ist die Verteilung beider Genfamilien in Insekten unregelmäßig und es wird angenommen, dass ihre Herkunft nicht auf einen anzestralen Ursprung zurückzuführen ist. Wahrscheinlicher ist es, dass sie separat erworben wurden - womöglich durch einen horizontalen Gentransfer (HGT) zwischen Mikroorganismus und Tier. Trotz der komplexen Evolution von GH45 und GH5_10 und der aufstrebenden Rolle von PZVE in der Biokraftstoffindustrie, sind beide Genfamilien in Phytophaga Käfern nur wenig untersucht. Das Hauptziel dieser These ist es daher PZVE der GH45 und GH5_10 aus Phytophagen Käfern -- mit Fokus auf enzymatischer Aktivität, physiologischer Bedeutung und Evolutionsgeschichte -- näher zu untersuchen.

(i) Eine funktionelle Charakterisierung der GH5_10 in *Gastrophysa viridula* (kodiert für eine Genkopie) und *Callosobruchus maculatus* (kodiert für vier Genkopien) zeigt, dass die ursprüngliche enzymatische Aktivität gegen Gluco- und Galactomannan erhalten blieb. Überraschenderweise hat eine der

GH5_10 aus *C. maculatus* seine mannanolytische Funktion verloren, erwarb aber stattdessen die Fähigkeit Xylan abzubauen. Weiterhin besitzt die GH5_10 Endo-β-1,4-mannanase aus *C. maculatus* die Fähigkeit zusätzlich Zellulose abzubauen, was möglicherweise eine Anpassung an das Fehlen von Zellulase-kodierenden Genfamilien in diesem Käfer darstellt. Eine Untersuchung der physiologischen Funktion der GH5_10 Endo-β-1,4-mannanase in *G. viridula* - mittels „Gen-Silencing“ durch RNA Interferenz (RNAi) - resultierte in einer signifikanten Herunterregulierung des Gens sowie in einer drastischen Reduzierung der korrespondierenden enzymatischen Aktivität. Es ist jedoch weder eine phänotypische Änderung in der Sterblichkeitsrate noch in der Gewichtzunahme zu beobachten, was wiederum eine komplexe Regulierung der GH5_10 und wahrscheinliche Kompensation durch andere PZVE (oder anderen Klassen von Verdauungsenzymen) nahe legt. Die Intron/Exon-Struktur von GH5_10 Genen in *G.viridula* und *C. maculatus* suggeriert, dass der letzte gemeinsame Vorfahr beider Arten bereits mindestens ein GH5_10 Gen besessen haben muss. Außerdem zeigt sich, dass evolutionsgeschichtlich nur Bacteria, Mollusca und Arthropoda GH5_10 in ihrem Genom kodieren. Interessanterweise fallen die GH5_10 der Insekten nicht in eine monophyletische Klade, sondern bilden zusammen mit GH5_10 anderer Arthropoden zwei eigenständige Gruppen. Dieses Ergebnis impliziert, dass GH5_10 Gene mindestens zweimal im Verlauf der Insektenevolution erworben wurden. Obwohl bisher kein Spenderorganismus identifiziert werden konnte, deuten die erhobenen Daten darauf hin, dass GH5_10 Gene in Insekten durch einen HGT erworben wurden.

(ii) *Gastrophysa viridula* wurde in dieser Arbeit als Modelorganismus verwendet um die enzymatischen Charakteristika sowie biologischen Funktionen von zwei GH45 zu ermitteln. Enzymatische Untersuchungen ermöglichen die Identifizierung einer GH45 Endo-β-1,4-glucanase, was mit dem Aktivitätsmuster von GH45 Enzymen aus Mikroben übereinstimmt. Interessanterweise hat die zweite GH45 ihre Aktivität gegenüber Zellulose verloren, hat im Verlauf der Evolution allerdings die Fähigkeit erworben

Xyloglucan abzubauen. Nach erfolgreicher Herunterregulierung beider GH45 Gene konnten allerdings keine Veränderungen bezüglich der Sterblichkeitsrate oder der Gewichtsänderung beobachtet werden. Basierend auf den Ergebnissen eines Zymographieexperiments wurden zusätzliche Enzyme mit zellulolytischer Aktivität in *G. viridula* festgestellt, welche höchstwahrscheinlich die herunterregulierte GH45 Zellulase enzymatisch ersetzen und einer Änderung des Phänotyps entgegenwirken. Im selben Experiment konnten keine zusätzlichen Xyloglucanasen in *G. viridula* identifiziert werden. Somit deuten die Ergebnisse auf eine ähnlich komplexe Regulation und Kompensation hin wie bereits für GH5_10 beobachtet. Diese Experimente zeigen deutlich, dass beide GH45 Gene aktive PZVE sind. Eines ist in der Lage amorphe Zellulose abzubauen, wohingegen das andere die Fähigkeit besitzt Xyloglucan abzubauen. Doch ähnlich wie schon für GH5_10 gezeigt, bleibt ihre biologische Rolle in Bezug auf die biologische Fitness der Käfer unklar.

(iii) Da GH45s nur selten in Käfern untersucht wurden und ihre funktionelle Charakterisierung sich fast ausschließlich auf Cerambycidae Käfer fokussiert, wurde die Untersuchung von GH45 auf vier zusätzliche Chrysomelidae sowie ein Curculionidae Käfer ausgeweitet. In Übereinstimmung mit den Ergebnissen aus *G. viridula* wurde in jeder Spezies mindestens eine GH45 Endo-β-1,4-Xyloglucanase und mindestens zwei GH45 Endo-β-1,4-Glucanases entdeckt. Darüber hinaus erkannten einige GH45 Endo-β-1,4-glucanases auch Glucomannan als Substrat. Eine phylogenetische Analyse der GH45 phytophager Käfer offenbarte, dass ihr letzter gemeinsame Vorfahr bereits mindestens eine GH45 Zellulase besaß. Zudem konnte gezeigt werden, dass der Substratwechsel von Zellulose zu Xyloglucan höchstwahrscheinlich durch eine Substitution von Aspartat zu Glutamat in katalytisch wichtigen Aminosäureresten hervorgerufen wurde. Allerdings erfolgte diese Substitution an zwei verschiedenen Aminosäurepositionen in Chrysomelidae und Curculionidae, was wiederum impliziert, dass die Xyloglucanaseaktivität unabhängig voneinander in beiden Käferfamilien entstand. Abschließend

wurde die Evolutionsgeschichte von GH45 in Tieren genauer untersucht. Es stellte sich hierbei heraus, dass GH45 in Arthropoden im Laufe der Evolution mindestens dreimal unabhängige voneinander erworben wurden. Die am nächsten zu den Käfer GH45 verwandte Klade beinhaltete Pilzsequenzen und suggeriert einen HGT der GH45 von Pilz zu Käfer. Zusammenfassend konnte gezeigt werden, dass GH45 Mitglieder innerhalb der Arthropoden mindestens dreimal unabhängig voneinander evolvierten und von Käfern vermutlich durch einen HGT -- mit Pilzen als Donororganismus -- erworben wurden. Nach diversen Genduplikationsereignissen und Substitutionen an katalytisch wichtigen Aminosäureresten, konnten einige GH45 Mitglieder neue Funktionen erwerben.

Diese Doktorarbeit hat wesentlich zum Verständnis der Funktion von Phytophaga Käfern kodierenden PZVE beigetragen. Dabei konnte unser Wissen über Käfer GH45 und GH5_10 maßgeblich vertieft werden. Die vorliegenden Ergebnisse zeigen, dass nicht ein vertikaler sondern ein horizontaler Gentransfer die mögliche Ursache der ursprünglichen Vererbung von GH45 (und wahrscheinlich GH5_10) in Käfern war. Die anschließenden spezies-spezifischen, unabhängigen Genduplikationen erlaubten eine funktionelle Diversifizierung und mögliche Adaption an die Nahrungsquelle der Käfer. Diese Untersuchungen geben einen fundamentalen Einblick in die Evolution von PZVE und der molekularen Mechanismen notwendig für den Erwerb neuer enzymatische Funktionen. Die Fähigkeit von GH45 und GH5_10 Pflanzenzellwandmaterial abzubauen, macht sie zu interessanten Kandidaten für die Anwendung in diversen Industriezweigen wie z.B. in Wein- und Bierbrauereien oder Textil- und Nahrungsmittelindustrie. Insbesondere können GH45 Zellulasen Anwendung in der Biokraftstoffproduktion finden und damit einem fortschreitenden Treibhauseffekt entgegenwirken.

7. References

Acuna, R, Padilla, BE, Florez-Ramos, CP, Rubio, JD, Herrera, JC, Benavides, P, Lee, SJ, Yeats, TH, Egan, AN, Doyle, JJ and Rose, JK (2012a) Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A* **109**: 4197-202.

Acuna, R, Padilla, BE, Florez-Ramos, CP, Rubio, JD, Herrera, JC, Benavides, P, Lee, SJ, Yeats, TH, Egan, AN, Doyle, JJ and Rose, JKC (2012b) Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 4197-4202.

Arimori, T, Ito, A, Nakazawa, M, Ueda, M and Tamada, T (2013) Crystal structure of endo-1,4-beta-glucanase from Eisenia fetida. *J Synchrotron Radiat* **20**: 884-9.

Aspeborg, H, Coutinho, PM, Wang, Y, Brumer, H, 3rd and Henrissat, B (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* **12**: 186.

Bayer, EA, Chanzy, H, Lamed, R and Shoham, Y (1998) Cellulose, cellulases and cellulosomes. *Curr Opin Struct Biol* **8**: 548-57.

Beguin, P and Aubert, JP (1994) The biological degradation of cellulose. *FEMS Microbiol Rev* **13**: 25-58.

Bhat, MK (2000) Cellulases and related enzymes in biotechnology. *Biotechnol Adv* **18**: 355-83.

Bolam, DN, Ciruela, A, McQueen-Mason, S, Simpson, P, Williamson, MP, Rixon, JE, Boraston, A, Hazlewood, GP and Gilbert, HJ (1998) Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* **331 (Pt 3)**: 775-81.

Boraston, AB, Bolam, DN, Gilbert, HJ and Davies, GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**: 769-81.

Boyle, PJ and Mitchell, R (1978) Absence of Microorganisms in Crustacean Digestive Tracts. *Science* **200**: 1157-1159.

Breznak, JA and Brune, A (1994) Role of Microorganisms in the Digestion of Lignocellulose by Termites. *Annual Review of Entomology* **39**: 453-487.

Brummell, DA and Maclachlan, GA (1989) Formation and Functions of Xyloglucan and Derivatives. *Acs Symposium Series* **399**: 18-32.

Brune, A (2014) Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* **12**: 168.

Buckeridge, MS, dos Santos, HP and Tine, MAS (2000) Mobilisation of storage cell wall polysaccharides in seeds. *Plant Physiology and Biochemistry* **38**: 141-156.

Bui, TH and Lee, SY (2015) Endogenous cellulase production in the leaf litter foraging mangrove crab Parasesarma erythodactyla. *Comp Biochem Physiol B Biochem Mol Biol* **179**: 27-36.

Busch, A, Kunert, G, Heckel, DG and 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**: e0184305.

Busch, A, Kunert, G, Wielsch, N and Pauchet, Y (2018a) Cellulose degradation in Gastrophysa viridula (Coleoptera: Chrysomelidae): functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity. *Insect Mol Biol* **27**: 633-650.

Busch, A, Kunert, G, Wielsch, N and Pauchet, Y (2018b) Cellulose degradation in Gastrophysa viridula (Coleoptera: Chrysomelidae): functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity. *Insect Mol Biol* **0**.

Busconi, M, Berzolla, A and Chiappini, E (2014) Preliminary data on cellulase encoding genes in the xylophagous beetle, Hylotrupes bajulus (Linnaeus). *International Biodeterioration & Biodegradation* **86**: 92-95.

Calderon-Cortes, N, Watanabe, H, Cano-Camacho, H, Zavala-Paramo, G and Quesada, M (2010) cDNA cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle Oncideres albomarginata chamela (Cerambycidae). *Insect Mol Biol* **19**: 323-36.

Capella-Gutierrez, S, Silla-Martinez, JM and Gabaldon, T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-3.

Chambost J.P., BMH, Cami B., Barras E. And Cattaneo J. (1987) Erwinia Cellulases. In: *Plant Pathogenic Bacteria*. pp. pp 150-159.

Chander Kuhad, R, Mehta, G, Gupta, R and Sharma, KK (2010) Fed batch enzymatic saccharification of newspaper cellulosics improves the sugar content in the hydrolysates and eventually the ethanol fermentation by Saccharomyces cerevisiae. *Biomass and Bioenergy* **34**: 1189-1194.

Chang, CJ, Wu, CP, Lu, SC, Chao, AL, Ho, TH, Yu, SM and Chao, YC (2012) A novel exo-cellulase from white spotted longhorn beetle (Anoplophora malasiaca). *Insect Biochem Mol Biol* **42**: 629-36.

Chang, MMC, T. Y. C.; Tsao, G. T. (1981) Structure, Pretreatment and Hydrolysis of Cellulose *Advances in Biochemical Engineering Bioenergy* **Vol. 20**: pp 15-42.

Chang, WH and Lai, AG (2018) Mixed evolutionary origins of endogenous biomass-depolymerizing enzymes in animals. *BMC Genomics* **19**: 483.

Colbourne, JK, Pfrender, ME, Gilbert, D, Thomas, WK, Tucker, A, Oakley, TH, Tokishita, S, Aerts, A, Arnold, GJ, Basu, MK, Bauer, DJ, Caceres, CE, Carmel, L, Casola, C, Choi, JH, Detter, JC, Dong, Q, Dusheyko, S, Eads, BD, Frohlich, T, Geiler-Samerotte, KA, Gerlach, D, Hatcher, P, Jogdeo, S, Krijgsveld, J, Kriventseva, EV, Kultz, D, Laforsch, C, Lindquist, E, Lopez, J, Manak, JR, Muller, J, Pangilinan, J, Patwardhan, RP, Pitluck, S, Pritham, EJ, Rechtsteiner, A, Rho, M, Rogozin, IB, Sakarya, O, Salamov, A, Schaack, S, Shapiro, H, Shiga, Y, Skalitzky, C, Smith, Z, Souvorov, A, Sung, W, Tang, Z, Tsuchiya, D, Tu, H, Vos, H, Wang, M, Wolf, YI, Yamagata, H, Yamada, T, Ye, Y, Shaw, JR, Andrews, J, Crease, TJ, Tang, H, Lucas, SM, Robertson, HM, Bork, P, Koonin, EV, Zdobnov, EM, Grigoriev, IV, Lynch, M and Boore, JL (2011) The ecoresponsive genome of Daphnia pulex. *Science* **331**: 555-61.

Cosgrove, DJ (1997) Assembly and enlargement of the primary cell wall in plants. *Annu Rev Cell Dev Biol* **13**: 171-201.

Cosgrove, DJ (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol* **6**: 850-61.

Cosgrove, DJ (2014) Re-constructing our models of cellulose and primary cell wall assembly. *Curr Opin Plant Biol* **22**: 122-31.

Crawley, M (2013) The R Book. *John Wiley and Sons Ldt.*

Danchin, EG, Rosso, MN, Vieira, P, de Almeida-Engler, J, Coutinho, PM, Henrissat, B and Abad, P (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A* **107**: 17651-6.

Davies, G and Henrissat, B (1995) Structures and Mechanisms of Glycosyl Hydrolases. *Structure* **3**: 853-859.

Davies, GJ, Tolley, SP, Henrissat, B, Hjort, C and Schulein, M (1995) Structures of oligosaccharide-bound forms of the endoglucanase V from Humicola insolens at 1.9 A resolution. *Biochemistry* **34**: 16210-20.

Davison, A and Blaxter, M (2005) Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol* **22**: 1273-84.

DeBoy, RT, Mongodin, EF, Fouts, DE, Tailford, LE, Khouri, H, Emerson, JB, Mohamoud, Y, Watkins, K, Henrissat, B, Gilbert, HJ and Nelson, KE (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. *J Bacteriol* **190**: 5455-63.

Dehal, P, Satou, Y, Campbell, RK, Chapman, J, Degnan, B, De Tomaso, A, Davidson, B, Di Gregorio, A, Gelpke, M, Goodstein, DM, Harafuji, N, Hastings, KEM, Ho, I, Hotta, K, Huang, W, Kawashima,

T, Lemaire, P, Martinez, D, Meinertzhagen, IA, Necula, S, Nonaka, M, Putnam, N, Rash, S, Saiga, H, Satake, M, Terry, A, Yamada, L, Wang, HG, Awazu, S, Azumi, K, Boore, J, Branno, M, Chin-bow, S, DeSantis, R, Doyle, S, Francino, P, Keys, DN, Haga, S, Hayashi, H, Hino, K, Imai, KS, Inaba, K, Kano, S, Kobayashi, K, Kobayashi, M, Lee, BI, Makabe, KW, Manohar, C, Matassi, G, Medina, M, Mochizuki, Y, Mount, S, Morishita, T, Miura, S, Nakayama, A, Nishizaka, S, Nomoto, H, Ohta, F, Oishi, K, Rigoutsos, I, Sano, M, Sasaki, A, Sasakura, Y, Shoguchi, E, Shin-i, T, Spagnuolo, A, Stainier, D, Suzuki, MM, Tassy, O, Takatori, N, Tokuoka, M, Yagi, K, Yoshizaki, F, Wada, S, Zhang, C, Hyatt, PD, Larimer, F, Detter, C, Doggett, N, Glavina, T, Hawkins, T, Richardson, P, Lucas, S, Kohara, Y, Levine, M, Satoh, N and Rokhsar, DS (2002) The draft genome of Ciona intestinalis: Insights into chordate and vertebrate origins. *Science* **298**: 2157-2167.

Dhiman, TR, Zaman, MS, Gimenez, RR, Walters, JL and Treacher, R (2002) Performance of dairy cows fed forage treated with fibrolytic enzymes prior to feeding. *Animal Feed Science and Technology* **101**: 115-125.

Dietrich, C, Kohler, T and Brune, A (2014) The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol* **80**: 2261-9.

Edwards, A (2002) Digestive enzymes of vine weevil (Otiorhynchus sulcatus) as potential targets for insect control strategies. In: *Book Digestive enzymes of vine weevil (Otiorhynchus sulcatus) as potential targets for insect control strategies* (Editor, ed.^eds.). Vol. Doctor of Philosophy, pp. Durham University, City.

Eyun, SI, Wang, H, Pauchet, Y, Ffrench-Constant, RH, Benson, AK, Valencia-Jimenez, A, Moriyama, EN and Siegfried, BD (2014) Molecular evolution of glycoside hydrolase genes in the Western corn rootworm (Diabrotica virgifera virgifera). *PLoS One* **9**: e94052.

Faddeeva-Vakhrusheva, A, Derks, MF, Anvar, SY, Agamennone, V, Suring, W, Smit, S, van Straalen, NM and Roelofs, D (2016) Gene Family Evolution Reflects Adaptation to Soil Environmental Stressors in the Genome of the Collembolan Orchesella cincta. *Genome Biol Evol* **8**: 2106-17.

Fischer, R, Ostafe, R and Twyman, RM (2013) Cellulases from insects. *Adv Biochem Eng Biotechnol* **136**: 51-64.

Fry, SC (1989) The Structure and Functions of Xyloglucan. *Journal of Experimental Botany* **40**: 1-11.

Fujita, K, Shimomura, K, Yamamoto, K, Yamashita, T and Suzuki, K (2006) A chitinase structurally related to the glycoside hydrolase family 48 is indispensable for the hormonally induced diapause termination in a beetle. *Biochem Biophys Res Commun* **345**: 502-7.

Gan, HM, Austin, C and Linton, S (2018) Transcriptome-Guided Identification of Carbohydrate Active Enzymes (CAZy) from the Christmas Island Red Crab, Gecarcoidea natalis and a Vote for the Inclusion of Transcriptome-Derived Crustacean CAZys in Comparative Studies. *Mar Biotechnol (NY)*.

Garvey, M, Klose, H, Fischer, R, Lambertz, C and Commandeur, U (2013) Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. *Trends in Biotechnology* **31**: 581-593.

Gaudin, C, Belaich, A, Champ, S and Belaich, JP (2000) CelE, a multidomain cellulase from Clostridium cellulolyticum: a key enzyme in the cellulosome? *J Bacteriol* **182**: 1910-5.

Gilbert, HJ (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol* **153**: 444-55.

Gilbert, HJ, Hall, J, Hazlewood, GP and Ferreira, LM (1990) The N-terminal region of an endoglucanase from Pseudomonas fluorescens subspecies cellulosa constitutes a cellulose-binding domain that is distinct from the catalytic centre. *Mol Microbiol* **4**: 759-67.

Girard, C and Jouanin, L (1999) Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, Phaedon cochleariae. *Insect Biochem Mol Biol* **29**: 1129-42.

Goyal, A, Ghosh, B and Eveleigh, D (1991) Characteristics of fungal cellulases. *Bioresource Technology* **36**: 37-50.

Guillen, D, Sanchez, S and Rodriguez-Sanoja, R (2010) Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol* **85**: 1241-9.

Guo, R, Ding, M, Zhang, SL, Xu, GJ and Zhao, FK (2008) Molecular cloning and characterization of two novel cellulase genes from the mollusc Ampullaria crossean. *J Comp Physiol B* **178**: 209-15.

Gusakov, AV (2013) Cellulases and hemicellulases in the 21st century race for cellulosic ethanol. *Biofuels* **4**: 567-569.

Hansen, AK and Moran, NA (2014) The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol* **23**: 1473-96.

Hayashi, T (1989) Xyloglucans in the Primary-Cell Wall. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**: 139-168.

Henrissat, B and Bairoch, A (1993) New Families in the Classification of Glycosyl Hydrolases Based on Amino-Acid-Sequence Similarities. *Biochemical Journal* **293**: 781-788.

Huang, Y, Niu, B, Gao, Y, Fu, L and Li, W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680-2.

Imjongjirak, C, Amparyup, P and Sittipraneed, S (2008) Cloning, genomic organization and expression of two glycosyl hydrolase family 10 (GHF10) genes from golden apple snail (Pomacea canaliculata). *DNA Seq* **19**: 224-36.

Irwin, DC, Zhang, S and Wilson, DB (2000) Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca. *Eur J Biochem* **267**: 4988-97.

Jacob, TA (1996) The effect of constant temperature and humidity on the development, longevity and productivity of Ahasverus advena (Waltl,) (Coleoptera: Silvanidae). *Journal of Stored Products Research* **32**: 115-121.

Javier, PFI, Óscar, G, Sanz-Aparicio, J and Díaz, P (2007) Xylanases: Molecular Properties and Applications. In: *Industrial Enzymes: Structure, Function and Applications* (Polaina, J and MacCabe, AP, eds.). pp. 65-82. Springer Netherlands, Dordrecht.

Joko, T, Subandi, A, Kusumandari, N, Wibowo, A and Priyatmojo, A (2014) Activities of plant cell wall-degrading enzymes by bacterial soft rot of orchid. *Archives of Phytopathology and Plant Protection* **47**: 1239-1250.

Kainulainen, P, Holopainen, J, Palomaki, V and Holopainen, T (1996) Effects of nitrogen fertilization on secondary chemistry and ectomycorrhizal state of Scots pine seedlings and on growth of grey pine aphid. *J Chem Ecol* **22**: 617-36.

Kapustin, Y, Souvorov, A, Tatusova, T and Lipman, D (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. *Biol Direct* **3**: 20.

Karim, N, Jones, JT, Okada, H and Kikuchi, T (2009) Analysis of expressed sequence tags and identification of genes encoding cell-wall-degrading enzymes from the fungivorous nematode Aphelenchus avenae. *BMC Genomics* **10**: 525.

Karmakar, M and Ray, RR (2011) Current trends in research and application of microbial cellulases. *Research Journal of Microbiology* **6**: 41-53.

Katoh, K and Standley, DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-80.

Keegstra, K (2010) Plant cell walls. *Plant Physiol* **154**: 483-6.

Keeling, CI, Yuen, MM, Liao, NY, Docking, TR, Chan, SK, Taylor, GA, Palmquist, DL, Jackman, SD, Nguyen, A, Li, M, Henderson, H, Janes, JK, Zhao, Y, Pandoh, P, Moore, R, Sperling, FA, Huber,

DP, Birol, I, Jones, SJ and Bohlmann, J (2013) Draft genome of the mountain pine beetle, Dendroctonus ponderosae Hopkins, a major forest pest. *Genome Biol* **14**: R27.

Kelkar, YD and Ochman, H (2012) Causes and Consequences of Genome Expansion in Fungi. *Genome Biology and Evolution* **4**: 13-23.

Kern, M, McGeehan, JE, Streeter, SD, Martin, RNA, Besser, K, Elias, L, Eborall, W, Malyon, GP, Payne, CM, Himmel, ME, Schnorr, K, Beckham, GT, Cragg, SM, Bruce, NC and McQueen-Mason, SJ (2013) Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 10189-10194.

Kerslake, JE, Woodin, SJ and Hartley, SE (1998) Effects of carbon dioxide and nitrogen enrichment on a plant-insect interaction: the quality of Calluna vulgaris as a host for Operophtera brumata. *New Phytologist* **140**: 43-53.

Kikuchi, T, Jones, JT, Aikawa, T, Kosaka, H and Ogura, N (2004) A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode Bursaphelenchus xylophilus. *FEBS Lett* **572**: 201-5.

Kim, N, Choo, YM, Lee, KS, Hong, SJ, Seol, KY, Je, YH, Sohn, HD and Jin, BR (2008) Molecular cloning and characterization of a glycosyl hydrolase family 9 cellulase distributed throughout the digestive tract of the cricket Teleogryllus emma. *Comp Biochem Physiol B Biochem Mol Biol* **150**: 368-76.

King, AJ, Cragg, SM, Li, Y, Dymond, J, Guille, MJ, Bowles, DJ, Bruce, NC, Graham, IA and 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 of the United States of America* **107**: 5345-5350.

Kirsch, R, Gramzow, L, Theissen, G, Siegfried, BD, Ffrench-Constant, RH, Heckel, DG and Pauchet, Y (2014) Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. *Insect Biochem Mol Biol* **52**: 33-50.

Kirsch, R, Heckel, DG and Pauchet, Y (2016) How the rice weevil breaks down the pectin network: Enzymatic synergism and sub-functionalization. *Insect Biochem Mol Biol* **71**: 72-82.

Kirsch, R, Wielsch, N, Vogel, H, Svatos, A, Heckel, DG and Pauchet, Y (2012) Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. *BMC Genomics* **13**: 587.

Knox, JP (2008) Revealing the structural and functional diversity of plant cell walls. *Curr Opin Plant Biol* **11**: 308-13.

Kobayashi, H, Nagahama, T, Arai, W, Sasagawa, Y, Umeda, M, Hayashi, T, Nikaido, I, Watanabe, H, Oguri, K, Kitazato, H, Fujioka, K, Kido, Y and Takami, H (2018) Polysaccharide hydrolase of the hadal zone amphipods Hirondellea gigas. *Biosci Biotechnol Biochem* **82**: 1123-1133.

Kostylev, M and Wilson, D (2012) Synergistic interactions in cellulose hydrolysis. *Biofules* **Volume 3**: Pages 61-70

Kostylev, M and Wilson, DB (2011) Determination of the catalytic base in family 48 glycosyl hydrolases. *Appl Environ Microbiol* **77**: 6274-6.

Kubicek, CP, Starr, TL and Glass, NL (2014) Plant Cell Wall–Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. *Annual Review of Phytopathology* **52**: 427-451.

Kuhad, RC, Gupta, R and Singh, A (2011a) Microbial cellulases and their industrial applications. *Enzyme Res* **2011**: 280696.

Kuhad, RC, Gupta, R and Singh, A (2011b) Microbial Cellulases and Their Industrial Applications. *Enzyme Research* **2011**: 10.

Kumar, S, Stecher, G and Tamura, K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870-4.

Kunii, M, Yasuno, M, Shindo, Y and Kawata, T (2014) A Dictyostelium cellobiohydrolase orthologue that affects developmental timing. *Development Genes and Evolution* **224**: 25-35.

Larsson, AM, Anderson, L, Xu, B, Munoz, IG, Uson, I, Janson, JC, Stalbrand, H and Stahlberg, J (2006) Three-dimensional crystal structure and enzymic characterization of beta-mannanase Man5A from blue mussel Mytilus edulis. *J Mol Biol* **357**: 1500-10.

Ledger, TN, Jaubert, S, Bosselut, N, Abad, P and Rosso, MN (2006) Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode Meloidogyne incognita and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. *Gene* **382**: 121-8.

Lee, SJ, Kim, SR, Yoon, HJ, Kim, I, Lee, KS, Je, YH, Lee, SM, Seo, SJ, Dae Sohn, H and Jin, BR (2004) cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, Apriona germari. *Comp Biochem Physiol B Biochem Mol Biol* **139**: 107-16.

Lee, SJ, Lee, KS, Kim, SR, Gui, ZZ, Kim, YS, Yoon, HJ, Kim, I, Kang, PD, Sohn, HD and Jin, BR (2005) A novel cellulase gene from the mulberry longicorn beetle, Apriona germari: gene structure, expression, and enzymatic activity. *Comp Biochem Physiol B Biochem Mol Biol* **140**: 551-60.

Leschine, SB (1995) Cellulose Degradation in Anaerobic Environments. *Annual Review of Microbiology* **49**: 399-426.

Li, QZ, Song, J, Peng, SB, Wang, JP, Qu, GZ, Sederoff, RR and Chiang, VL (2014) Plant biotechnology for lignocellulosic biofuel production. *Plant Biotechnology Journal* **12**: 1174-1192.

Li, Y, Yin, Q, Ding, M and Zhao, F (2009) Purification, characterization and molecular cloning of a novel endo-beta-1,4-glucanase AC-EG65 from the mollusc Ampullariacrossean. *Comp Biochem Physiol B Biochem Mol Biol* **153**: 149-56.

Libertini, E, Li, Y and McQueen-Mason, SJ (2004) Phylogenetic analysis of the plant endo-beta-1,4-glucanase gene family. *J Mol Evol* **58**: 506-15.

Liu, G, Wei, X, Qin, Y and Qu, Y (2010) Characterization of the endoglucanase and glucomannanase activities of a glycoside hydrolase family 45 protein from Penicillium decumbens 114-2. *J Gen Appl Microbiol* **56**: 223-9.

Liu, Y-S, Baker, JO, Zeng, Y, Himmel, ME, Haas, T and Ding, S-Y (2011) Cellobiohydrolase Hydrolyzes Crystalline Cellulose on Hydrophobic Faces. *The Journal of Biological Chemistry* **286**: 11195-11201.

Lombard, V, Golaconda Ramulu, H, Drula, E, Coutinho, PM and Henrissat, B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490-5.

Ma, Z, Zhu, L, Song, T, Wang, Y, Zhang, Q, Xia, Y, Qiu, M, Lin, Y, Li, H, Kong, L, Fang, Y, Ye, W, Wang, Y, Dong, S, Zheng, X, Tyler, BM and Wang, Y (2017) A paralogous decoy protects Phytophthora sojae apoplastic effector PsXEG1 from a host inhibitor. *Science* **355**: 710-714.

Majka, CG (2007) The Ciidae (Coleoptera : Tenebrionoidea) of the Maritime Provinces of Canada: new records, distribution, zoogeography, and observations on beetle-fungi relationships in saproxylic environments. *Zootaxa*: 1-20.

Mansfield, SD, Mooney, C and Saddler, JN (1999) Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnology Progress* **15**: 804-816.

Marvaldi, AE, Duckett, CN, Kjer, KM and Gillespie, JJ (2009) Structural alignment of 18S and 28S rDNA sequences provides insights into phylogeny of Phytophaga (Coleoptera: Curculionoidea and Chrysomeloidea). *Zoologica Scripta* **38**: 63-77.

Matano, Y, Hasunuma, T and Kondo, A (2013) Cell recycle batch fermentation of high-solid lignocellulose using a recombinant cellulase-displaying yeast strain for high yield ethanol production in consolidated bioprocessing. *Bioresource Technology* **135**: 403-409.

Matte, A, Kozlov, G, Trempe, J-F, Currie, MA, Burk, D, Jia, Z, Gehring, K, Ekiel, I, Berghuis, AM and Cygler, M (2009) Preparation and Characterization of Bacterial Protein Complexes for

Structural Analysis. In: *Advances in Protein Chemistry and Structural Biology* (Joachimiak, A, ed. Vol. 76, pp. 1-42. Academic Press.

Matthysse, AG, Deschet, K, Williams, M, Marry, M, White, AR and Smith, WC (2004) A functional cellulose synthase from ascidian epidermis. *Proc Natl Acad Sci U S A* **101**: 986-91.

Mcgavin, M and Forsberg, CW (1988) Isolation and Characterization of Endoglucanase-1 and Endoglucanase-2 from Bacteroides-Succinogenes S85. *Journal of Bacteriology* **170**: 2914-2922.

McKenna, DD, Scully, ED, Pauchet, Y, Hoover, K, Kirsch, R, Geib, SM, Mitchell, RF, Waterhouse, RM, Ahn, SJ, Arsala, D, Benoit, JB, Blackmon, H, Bledsoe, T, Bowsher, JH, Busch, A, Calla, B, Chao, H, Childers, AK, Childers, C, Clarke, DJ, Cohen, L, Demuth, JP, Dinh, H, Doddapaneni, H, Dolan, A, Duan, JJ, Dugan, S, Friedrich, M, Glastad, KM, Goodisman, MA, Haddad, S, Han, Y, Hughes, DS, Ioannidis, P, Johnston, JS, Jones, JW, Kuhn, LA, Lance, DR, Lee, CY, Lee, SL, Lin, H, Lynch, JA, Moczek, AP, Murali, SC, Muzny, DM, Nelson, DR, Palli, SR, Panfilio, KA, Pers, D, Poelchau, MF, Quan, H, Qu, J, Ray, AM, Rinehart, JP, Robertson, HM, Roehrdanz, R, Rosendale, AJ, Shin, S, Silva, C, Torson, AS, Jentzsch, IM, Werren, JH, Worley, KC, Yocum, G, Zdobnov, EM, Gibbs, RA and Richards, S (2016) Genome of the Asian longhorned beetle (Anoplophora glabripennis), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. *Genome Biol* **17**: 227.

Mei, HZ, Xia, DG, Zhao, QL, Zhang, GZ, Qiu, ZY, Qian, P and Lu, C (2015) Molecular cloning, expression, purification and characterization of a novel cellulase gene (Bh-EGaseI) in the beetle Batocera horsfieldi. *Gene*.

Mei, HZ, Xia, DG, Zhao, QL, Zhang, GZ, Qiu, ZY, Qian, P and Lu, C (2016) Molecular cloning, expression, purification and characterization of a novel cellulase gene (Bh-EGaseI) in the beetle Batocera horsfieldi. *Gene* **576**: 45-51.

Meier, H and Reid, JSG (1982) Reserve Polysaccharides Other Than Starch in Higher Plants. In: *Plant Carbohydrates I: Intracellular Carbohydrates* (Loewus, FA and Tanner, W, eds.). pp. 418-471. Springer Berlin Heidelberg, Berlin, Heidelberg.

Montanari, S, Rountani, M, Heux, L and Vignon, MR (2005) Topochemistry of carboxylated cellulose nanocrystals resulting from TEMPO-mediated oxidation. *Macromolecules* **38**: 1665-1671.

Mueller, UG, Gerardo, NM, Aanen, DK, Six, DL and Schultz, TR (2005) The Evolution of Agriculture in Insects. *Annual Review of Ecology, Evolution, and Systematics* **36**: 563-595.

Nagy, LG, Riley, R, Tritt, A, Adam, C, Daum, C, Floudas, D, Sun, H, Yadav, JS, Pangilinan, J, Larsson, KH, Matsuura, K, Barry, K, Labutti, K, Kuo, R, Ohm, RA, Bhattacharya, SS, Shirouzu, T, Yoshinaga, Y, Martin, FM, Grigoriev, IV and Hibbett, DS (2016) Comparative Genomics of Early-Diverging Mushroom-Forming Fungi Provides Insights into the Origins of Lignocellulose Decay Capabilities. *Mol Biol Evol* **33**: 959-70.

Nakashima, K, Yamada, L, Satou, Y, Azuma, J and Satoh, N (2004) The evolutionary origin of animal cellulose synthase. *Dev Genes Evol* **214**: 81-8.

Neuhoff, V, Stamm, R and Eibl, H (1985) Clear Background and Highly Sensitive Protein Staining with Coomassie Blue Dyes in Polyacrylamide Gels - a Systematic Analysis. *Electrophoresis* **6**: 427-448.

Nguyen, STC, Freund, HL, Kasanjian, J and Berlemont, R (2018) Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. *Applied Microbiology and Biotechnology* **102**: 1629-1637.

Ni, J and Tokuda, G (2013) Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol Adv* **31**: 838-50.

Nishida, Y, Suzuki, K, Kumagai, Y, Tanaka, H, Inoue, A and Ojima, T (2007) Isolation and primary structure of a cellulase from the Japanese sea urchin Strongylocentrotus nudus. *Biochimie* **89**: 1002-11.

O'Connor, RM, Fung, JM, Sharp, KH, Benner, JS, McClung, C, Cushing, S, Lamkin, ER, Fomenkov, AI, Henrissat, B, Londer, YY, Scholz, MB, Posfai, J, Malfatti, S, Tringe, SG, Woyke, T, Malmstrom, RR, Coleman-Derr, D, Altamia, MA, Dedrick, S, Kaluziak, ST, Haygood, MG and Distel, DL (2014) Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. *Proc Natl Acad Sci U S A* **111**: E5096-104.

Obeng, EM, Adam, SNN, Budiman, C, Ongkudon, CM, Maas, R and Jose, J (2017) Lignocellulases: a review of emerging and developing enzymes, systems, and practices. *Bioresources and Bioprocessing* **4**: 16.

Ootsuka, S, Saga, N, Suzuki, K, Inoue, A and Ojima, T (2006) Isolation and cloning of an endo-beta-1,4-mannanase from Pacific abalone Haliotis discus hannai. *J Biotechnol* **125**: 269-80.

Padilla-Hurtado, B, Florez-Ramos, C, Aguilera-Galvez, C, Medina-Olaya, J, Ramirez-Sanjuan, A, Rubio-Gomez, J and Acuna-Zornosa, R (2012) Cloning and expression of an endo-1,4-beta-xylanase from the coffee berry borer, Hypothenemus hampei. *BMC Res Notes* **5**: 23.

Palomares-Rius, JE, Hirooka, Y, Tsai, IJ, Masuya, H, Hino, A, Kanzaki, N, Jones, JT and Kikuchi, T (2014) Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. *BMC Evol Biol* **14**: 69.

Parisutham, V, Kim, TH and Lee, SK (2014) Feasibilities of consolidated bioprocessing microbes: From pretreatment to biofuel production. *Bioresource Technology* **161**: 431-440.

Parsiegla, G, Juy, M, Reverbel-Leroy, C, Tardif, C, Belaich, JP, Driguez, H and Haser, R (1998) The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 A resolution. *Embo j* **17**: 5551-62.

Pauchet, Y and Heckel, DG (2013) The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer. *Proc Biol Sci* **280**: 20131021.

Pauchet, Y, Kirsch, R, Giraud, S, Vogel, H and Heckel, DG (2014a) Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle Apriona japonica. *Insect Biochem Mol Biol* **49**: 1-13.

Pauchet, Y, Saski, CA, Feltus, FA, Luyten, I, Quesneville, H and Heckel, DG (2014b) Studying the organization of genes encoding plant cell wall degrading enzymes in *Chrysomela tremula* provides insights into a leaf beetle genome. *Insect Mol Biol* **23**: 286-300.

Pauchet, Y, Wilkinson, P, Chauhan, R and Ffrench-Constant, RH (2010) Diversity of Beetle Genes Encoding Novel Plant Cell Wall Degrading Enzymes. *Plos One* **5**: e15635.

Pauly, M, Gille, S, Liu, L, Mansoori, N, de Souza, A, Schultink, A and Xiong, G (2013) Hemicellulose biosynthesis. *Planta* **238**: 627-42.

Pauly, M and Keegstra, K (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* **54**: 559-68.

Payne, CM, Knott, BC, Mayes, HB, Hansson, H, Himmel, ME, Sandgren, M, Stahlberg, J and Beckham, GT (2015) Fungal cellulases. *Chem Rev* **115**: 1308-448.

Pere, J, Puolakka, A, Nousiainen, P and Buchert, J (2001) Action of purified Trichoderma reesei cellulases on cotton fibers and yarn. *Journal of Biotechnology* **89**: 247-255.

Pinheiro, J, Bates, D, DebRoy, S, Sarkar, D and Team, RC (2015) Linear and Nonlinear Mixed Effects Models. *http://CRAN.R-project.org/package=nlme*.

Py, B, Bortoli-German, I, Haiech, J, Chippaux, M and Barras, F (1991) Cellulase EGZ of Erwinia chrysanthemi: structural organization and importance of His98 and Glu133 residues for catalysis. *Protein Eng* **4**: 325-33.

R-Development-Core-Team (2015) R: A language and environment for statistical computing. *Vienna, Austria: R Foundation for Statistical Computing*.

Rahman, MM, Inoue, A and Ojima, T (2014) Characterization of a GHF45 cellulase, AkEG21, from the common sea hare Aplysia kurodai. *Frontiers in Chemistry* **2**.

Reverbel-Leroy, C, Pages, S, Belaich, A, Belaich, JP and Tardif, C (1997) The processive endocellulase CelF, a major component of the Clostridium cellulolyticum cellulosome: purification and characterization of the recombinant form. *J Bacteriol* **179**: 46-52.

Rincon, MT, McCrae, SI, Kirby, J, Scott, KP and Flint, HJ (2001) EndB, a multidomain family 44 cellulase from Ruminococcus flavefaciens 17, binds to cellulose via a novel cellulose-binding module and to another R. flavefaciens protein via a dockerin domain. *Appl Environ Microbiol* **67**: 4426-31.

Ronquist, F, Teslenko, M, van der Mark, P, Ayres, DL, Darling, A, Hohna, S, Larget, B, Liu, L, Suchard, MA and Huelsenbeck, JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539-42.

Rossi, AM, Brodbeck, BV and Strong, DR (1996) Response of xylem-feeding leafhopper to host plant species and plant quality. *J Chem Ecol* **22**: 653-71.

Ruel, K, Nishiyama, Y and Joseleau, JP (2012) Crystalline and amorphous cellulose in the secondary walls of Arabidopsis. *Plant Sci* **193-194**: 48-61.

Russell, JA, Moreau, CS, Goldman-Huertas, B, Fujiwara, M, Lohman, DJ and Pierce, NE (2009) Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 21236-21241.

Sakamoto, K and Toyohara, H (2009) Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam Corbicula japonica. *Comp Biochem Physiol B Biochem Mol Biol* **152**: 390-6.

Saleh, F, Tahir, M, Ohtsuka, A and Hayashi, K (2005) A mixture of pure cellulase, hemicellulase and pectinase improves broiler performance. *Br Poult Sci* **46**: 602-6.

Salem, H, Bauer, E, Kirsch, R, Berasategui, A, Cripps, M, Weiss, B, Koga, R, Fukumori, K, Vogel, H, Fukatsu, T and Kaltenpoth, M (2017) Drastic Genome Reduction in an Herbivore's Pectinolytic Symbiont. *Cell* **171**: 1520-1531 e13.

Saloheimo, A, Henrissat, B, Hoffren, AM, Teleman, O and Penttila, M (1994) A novel, small endoglucanase gene, egl5, from Trichoderma reesei isolated by expression in yeast. *Mol Microbiol* **13**: 219-28.

Sanchez, MM, Pastor, FI and Diaz, P (2003) Exo-mode of action of cellobiohydrolase Cel48C from Paenibacillus sp. BP-23. A unique type of cellulase among Bacillales. *Eur J Biochem* **270**: 2913-9.

Saxena, IMaB, R.M. (2007) A perspective on the assembly of cellulose-synthesizing complexes: Possible role of korrigan and microtubules in cellulose synthesis in plants. *Cellulose: Molecular and Structural Biology*: pp 169-181.

Scheller, HV and Ulvskov, P (2010) Hemicelluloses. *Annual Review of Plant Biology, Vol 61* **61**: 263-289.

Schigel, DS (2012) Fungivory and host associations of Coleoptera: a bibliography and review of research approaches. *Mycology* **3**: 258-272.

Schoville, SD, Chen, YH, Andersson, MN, Benoit, JB, Bhandari, A, Bowsher, JH, Brevik, K, Cappelle, K, Chen, MM, Childers, AK, Childers, C, Christiaens, O, Clements, J, Didion, EM, Elpidina, EN, Engsontia, P, Friedrich, M, Garcia-Robles, I, Gibbs, RA, Goswami, C, Grapputo, A, Gruden, K, Grynberg, M, Henrissat, B, Jennings, EC, Jones, JW, Kalsi, M, Khan, SA, Kumar, A, Li, F, Lombard, V, Ma, X, Martynov, A, Miller, NJ, Mitchell, RF, Munoz-Torres, M, Muszewska, A, Oppert, B, Palli, SR, Panfilio, KA, Pauchet, Y, Perkin, LC, Petek, M, Poelchau, MF, Record, E,

Rinehart, JP, Robertson, HM, Rosendale, AJ, Ruiz-Arroyo, VM, Smagghe, G, Szendrei, Z, Thomas, GWC, Torson, AS, Vargas Jentzsch, IM, Weirauch, MT, Yates, AD, Yocum, GD, Yoon, JS and Richards, S (2018) A model species for agricultural pest genomics: the genome of the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). *Sci Rep* **8**: 1931.

Schulein, M (1997) Enzymatic properties of cellulases from Humicola insolens. *Journal of Biotechnology* **57**: 71-81.

Schulein, M (1998) Kinetics of fungal cellulases. *Biochem Soc Trans* **26**: 164-7.

Schultink, A, Liu, L, Zhu, L and Pauly, M (2014) Structural Diversity and Function of Xyloglucan Sidechain Substituents. *Plants (Basel)* **3**: 526-42.

Shelomi, M, Heckel, DG and 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 and Arakawa, G (2014) Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut. *J Insect Physiol* **60**: 25-30.

Shen, H, Gilkes, NR, Kilburn, DG, Miller, RC and Warren, RAJ (1995) Cellobiohydrolase-B, a 2nd Exo-Cellobiohydrolase from the Cellulolytic Bacterium Cellulomonas-Fimi. *Biochemical Journal* **311**: 67-74.

Sheppard, PO, Grant, FJ, Oort, PJ, Sprecher, CA, Foster, DC, Hagen, FS, Upshall, A, McKnight, GL and O'Hara, PJ (1994) The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from Fusarium oxysporum. *Gene* **150**: 163-7.

Shevchenko, A, Sunyaev, S, Loboda, A, Shevehenko, A, Bork, P, Ens, W and Standing, KG (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. *Analytical Chemistry* **73**: 1917-1926.

Shevchenko, A, Tomas, H, Havlis, J, Olsen, JV and Mann, M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**: 2856-60.

Shrivastava, B, Thakur, S, Khasa, YP, Gupte, A, Puniya, AK and Kuhad, RC (2011) White-rot fungal conversion of wheat straw to energy rich cattle feed. *Biodegradation* **22**: 823-831.

Silva, GA and Lapenta, AS (2011) Genetic variability in esterases and the insecticide resistance in brazilian strains of Oryzaephilus mercator and Oryzaephilus surinamensis (Coleoptera: Silvanidae). *Bull Entomol Res* **101**: 177-85.

Sims, REH, Mabee, W, Saddler, JN and Taylor, M (2010) An overview of second generation biofuel technologies. *Bioresource Technology* **101**: 1570-1580.

Sinha, RN (1959a) THE HYDROGEN-ION CONCENTRATION IN THE ALIMENTARY CANAL OF BEETLES INFESTING STORED GRAIN AND GRAIN PRODUCTS. *Annals of the Entomological Society of America*.

Sinha, RN (1959b) The Hydrogen-Ion Concentration in the Alimentary Canal of Beetles Infesting Stored Grain and Grain Products1. *Annals of the Entomological Society of America* **52**: 763-765.

Slaytor, M (1992) Cellulose digestion in termites and cockroaches: What role do symbionts play? *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **103**: 775-784.

Smant, G, Stokkermans, JP, Yan, Y, de Boer, JM, Baum, TJ, Wang, X, Hussey, RS, Gommers, FJ, Henrissat, B, Davis, EL, Helder, J, Schots, A and Bakker, J (1998) Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci U S A* **95**: 4906-11.

Somerville, C (2006) Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol* **22**: 53-78.

Song, JM, Hong, SK, An, YJ, Kang, MH, Hong, KH, Lee, YH and Cha, SS (2017) Genetic and Structural Characterization of a Thermo-Tolerant, Cold-Active, and Acidic Endo-beta-1,4-glucanase from Antarctic Springtail, Cryptopygus antarcticus. *J Agric Food Chem* **65**: 1630-1640.

Song, JM, Nam, KW, Kang, SG, Kim, CG, Kwon, ST and Lee, YH (2008) Molecular cloning and characterization of a novel cold-active beta-1,4-D-mannanase from the Antarctic springtail, Cryptopygus antarcticus. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **151**: 32-40.

Stamatakis, A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-3.

Stern, J, Kahn, A, Vazana, Y, Shamshoum, M, Morais, S, Lamed, R and Bayer, EA (2015) Significance of Relative Position of Cellulases in Designer Cellulosomes for Optimized Cellulolysis. *Plos One* **10**.

Sugimura, M, Watanabe, H, Lo, N and Saito, H (2003) Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, Psacothea hilaris. *European Journal of Biochemistry* **270**: 3455-3460.

Sukharnikov, LO, Alahuhta, M, Brunecky, R, Upadhyay, A, Himmel, ME, Lunin, VV and Zhulin, IB (2012) Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. *J Biol Chem* **287**: 41068-77.

Suzuki, K, Ojima, T and Nishita, K (2003) Purification and cDNA cloning of a cellulase from abalone Haliotis discus hannai. *Eur J Biochem* **270**: 771-8.

Szydlowski, L, Boschetti, C, Crisp, A, Barbosa, EG and Tunnacliffe, A (2015) Multiple horizontally acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the bdelloid rotifer Adineta ricciae. *Gene* **566**: 125-37.

Takahashi, M, Takahashi, H, Nakano, Y, Konishi, T, Terauchi, R and Takeda, T (2010) Characterization of a cellobiohydrolase (MoCel6A) produced by Magnaporthe oryzae. *Appl Environ Microbiol* **76**: 6583-90.

Tan, MH, Gan, HM, Gan, HY, Lee, YP, Croft, LJ, Schultz, MB, Miller, AD and Austin, CM (2016) First comprehensive multi-tissue transcriptome of Cherax quadricarinatus (Decapoda: Parastacidae) reveals unexpected diversity of endogenous cellulase. *Organisms Diversity & Evolution* **16**: 185-200.

Tanabe, T, Morinaga, K, Fukamizo, T and Mitsutomi, M (2003) Novel chitosanase from Streptomyces griseus HUT 6037 with transglycosylation activity. *Biosci Biotechnol Biochem* **67**: 354-64.

Thomas, LH, Forsyth, VT, Sturcova, A, Kennedy, CJ, May, RP, Altaner, CM, Apperley, DC, Wess, TJ and Jarvis, MC (2013) Structure of cellulose microfibrils in primary cell walls from collenchyma. *Plant Physiol* **161**: 465-76.

Treangen, TJ and Rocha, EP (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7**: e1001284.

Tsuji, A, Tominaga, K, Nishiyama, N and Yuasa, K (2013) Comprehensive enzymatic analysis of the cellulolytic system in digestive fluid of the Sea Hare Aplysia kurodai. Efficient glucose release from sea lettuce by synergistic action of 45 kDa endoglucanase and 210 kDa ss-glucosidase. *PLoS One* **8**: e65418.

Ueda, M, Hirano, Y, Fukuhara, H, Naka, Y, Nakazawa, M, Sakamoto, T, Ogata, Y and Tamada, T (2018) Gene cloning, expression, and X-ray crystallographic analysis of a beta-mannanase from Eisenia fetida. *Enzyme Microb Technol* **117**: 15-22.

Valencia, A, Alves, AP and 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**: 260-7.

Valencia Jimenez, A, Wang, H and Siegfried, BD (2014) Expression and characterization of a recombinant endoglucanase from western corn rootworm, in Pichia pastoris. *J Insect Sci* **14**: 242.

Vazana, Y, Morais, S, Barak, Y, Lamed, R and Bayer, EA (2012) Designer cellulosomes for enhanced hydrolysis of cellulosic substrates. *Methods Enzymol* **510**: 429-52.

Vega, FE, Brown, SM, Chen, H, Shen, E, Nair, MB, Ceja-Navarro, JA, Brodie, EL, Infante, F, Dowd, PF and Pain, A (2015) Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer, Hypothenemus hampei. *Sci Rep* **5**: 12525.

Viëtor, RJ, Newman, RH, Ha, M-A, Apperley, DC and Jarvis, MC (2002) Conformational features of crystal-surface cellulose from higher plants. *The Plant Journal* **30**: 721-731.

Wagner, A (2008) Gene duplications, robustness and evolutionary innovations. *Bioessays* **30**: 367-373.

Wang, J, Ding, M, Li, YH, Chen, QX, Xu, GJ and Zhao, FK (2003) Isolation of a multi-functional endogenous cellulase gene from mollusc, Ampullaria crossean. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **35**: 941-6.

Wang, JC, Zeng, QW, Zhou, LF and Chen, FM (2018) Molecular characteristics and functional analysis of the beta-1,4-endoglucanase Bm-eng-1 gene of Bursaphelenchus mucronatus (Nematoda: Aphelenchoididae). *Forest Pathology* **48**.

Wang, S, Zhang, J, Jiao, W, Li, J, Xun, X, Sun, Y, Guo, X, Huan, P, Dong, B, Zhang, L, Hu, X, Sun, X, Wang, J, Zhao, C, Wang, Y, Wang, D, Huang, X, Wang, R, Lv, J, Li, Y, Zhang, Z, Liu, B, Lu, W, Hui, Y, Liang, J, Zhou, Z, Hou, R, Li, X, Liu, Y, Li, H, Ning, X, Lin, Y, Zhao, L, Xing, Q, Dou, J, Li, Y, Mao, J, Guo, H, Dou, H, Li, T, Mu, C, Jiang, W, Fu, Q, Fu, X, Miao, Y, Liu, J, Yu, Q, Li, R, Liao, H, Li, X, Kong, Y, Jiang, Z, Chourrout, D, Li, R and Bao, Z (2017) Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol* **1**: 120.

Watanabe, H, Noda, H, Tokuda, G and Lo, N (1998) A cellulase gene of termite origin. *Nature* **394**: 330-1.

Watanabe, H and Tokuda, G (2010) Cellulolytic systems in insects. *Annu Rev Entomol* **55**: 609-32.

Willis, JD, Oppert, B, Oppert, C, Klingeman, WE and Jurat-Fuentes, JL (2011) Identification, cloning, and expression of a GHF9 cellulase from Tribolium castaneum (Coleoptera: Tenebrionidae). *J Insect Physiol* **57**: 300-6.

Wilson, DB (2009) Cellulases and biofuels. *Current Opinion in Biotechnology* **20**: 295-299.

Xia, D, Wei, Y, Zhang, G, Zhao, Q, Zhang, Y, Xiang, Z and Lu, C (2013) cDNA cloning, expression, and enzymatic activity of a novel endogenous cellulase from the beetle Batocera horsfieldi. *Gene* **514**: 62-8.

Xu, B, Janson, JC and Sellos, D (2001) Cloning and sequencing of a molluscan endo-beta-1,4-glucanase gene from the blue mussel, Mytilus edulis. *Eur J Biochem* **268**: 3718-27.

Xu, BZ, Hagglund, P, Stalbrand, H and Janson, JC (2002) endo-beta-1,4-mannanases from blue mussel, Mytilus edulis: purification, characterization, and mode of action. *Journal of Biotechnology* **92**: 267-277.

Zahura, UA, Rahman, MM, Inoue, A, Tanaka, H and Ojima, T (2011) cDNA cloning and bacterial expression of an endo-beta-1,4-mannanase, AkMan, from Aplysia kurodai. *Comp Biochem Physiol B Biochem Mol Biol* **159**: 227-35.

Zhang, G, Fang, X, Guo, X, Li, L, Luo, R, Xu, F, Yang, P, Zhang, L, Wang, X, Qi, H, Xiong, Z, Que, H, Xie, Y, Holland, PW, Paps, J, Zhu, Y, Wu, F, Chen, Y, Wang, J, Peng, C, Meng, J, Yang, L, Liu, J, Wen, B, Zhang, N, Huang, Z, Zhu, Q, Feng, Y, Mount, A, Hedgecock, D, Xu, Z, Liu, Y, Domazet-Loso, T, Du, Y, Sun, X, Zhang, S, Liu, B, Cheng, P, Jiang, X, Li, J, Fan, D, Wang, W, Fu, W, Wang, T, Wang, B, Zhang, J, Peng, Z, Li, Y, Li, N, Wang, J, Chen, M, He, Y, Tan, F, Song, X, Zheng, Q, Huang, R, Yang, H, Du, X, Chen, L, Yang, M, Gaffney, PM, Wang, S, Luo, L, She, Z, Ming, Y, Huang, W, Zhang, S, Huang, B, Zhang, Y, Qu, T, Ni, P, Miao, G, Wang, J, Wang, Q, Steinberg, CE, Wang, H,

Li, N, Qian, L, Zhang, G, Li, Y, Yang, H, Liu, X, Wang, J, Yin, Y and Wang, J (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**: 49-54.

Zhang, YH, Cui, J, Lynd, LR and Kuang, LR (2006) A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7**: 644-8.

Zuur, A, Fleno, EN, Walker, N, Saveliev, A and Smith, GM (2009) Mixed effects models and extensions in ecology with R.

8. Danksagung

An allererster Stelle möchte ich Yannick danken, der nie aufgegeben hat aus mir einen vernünftigen Wissenschaftler zu machen. Vielen Dank für die Geduld bei der Vorbereitung von Vorträgen oder wenn's einfach mal wieder länger dauerte, für den Sachverstand in verzwickten Angelegenheiten, den Einfallsreichtum und für die motivierenden Worte. Man kann nur schwer einen besseren Betreuer bekommen.

Ein großer Dank geht auch an David G. Heckel, der mir die Möglichkeit gab an diesem Thema zu arbeiten und mir mit seinem erstaunlichen Repertoire an Wissen nicht nur neue Einblicke in die Welt der Insekten bescherte, sondern mir auch half einige meiner experimentellen Rückschlüsse nochmal ganz neu zu überdenken.

Außerdem möchte ich der ganzen „Digestion-Group" danken; also Roy, Wiebke, Pauline, Nara und Bianca, die mir immer mit Rat und Tat zur Seite standen, wussten wo was zu finden war oder einfach nur die gleiche Musik im Labor mochten. Besonders sei hier noch Roy erwähnt, mit dem ich oft die Ergebnisse meiner Arbeit diskutiert habe, und der mir half sie unter dem richtigen Blickwinkel zu betrachten.

Vielen Dank auch an Grit für ihre Hilfe und Antwort bei statistischen Problemen sowie Natalie und Yvonne für ihre Einführung in die Welt der Massenspektrometrie.

Danke an meine Bürokumpanen David, Anne, Shantanu und Pauline, welche immer für eine kleine Runde „trash talk" zu haben waren und zu einer tollen Büroatmosphäre beitrugen. Außerdem danke, dass ihr euren Senf immer dazu beigetragen habt wenn ich Fragen hatte (oder auch wenn ich keine hatte). Dankschön nochmal im Speziellen an Shantanu und David für die technische Unterstützung in Sachen Statistik, RNAi und qPCR.

Ich möchte mich auch bei meiner Bachelorstudentin Sina bedanken, die mir mit ihren im Rahmen ihrer Abschlussarbeit erhobenen Daten eine große Hilfe war und mir zeigte welche Verantwortung und gleichzeitig Spaß die wissenschaftliche Betreuung von Studenten mit sich bringt.

Und wenn ich mal wieder was nicht gefunden habe oder Unterstützung im Labor brauchte, konnte ich mich auch immer auf unsere technischen Assistenten verlassen. Danke Susi und Antje, Jette und Dominika, Regina und Angelika. Angelika möchte ich besonders erwähnen, die sich unermüdlich und mit großer Hingabe um „meine Lieblinge" kümmerte und von dem

Schicksal mancher auch lieber gar nichts wissen wollte. Steffi, unsere Bioinformatikerin, die viel Zeit in die Etablierung einiger meiner Experimente miteingebracht hat, soll natürlich auch nicht unerwähnt bleiben.

Vieles, was diese Doktorarbeit ausmachte, wurde auch im Labor oder gleich auf den Gängen unserer Abteilung besprochen. Zum Beispiel wenn es um experimentelles Design, Methodenprobleme, Posteraufbau oder Vortragskonzepte ging. Gefühlt hatte deshalb irgendwie jeder der Abteilung seinen Beitrag zu meiner Doktorarbeit geleistet, daher vielen Dank an den Rest des E-Teams (Theresa, Melanie, Andrea, Astrid, Franzi, Zhi Ling, Suyog, Matilda, Silvia, Heiko, Yu, Sher, Sebastian, Sabine, Chris und Hanna).

Einen großen Dank auch an die Haustechnik und die Gewächshauscrew, die immer für diverse Projekte hilfsbereit zur Seite standen. Hier sei auch IT, Bibliothek und Buchhaltung genannt, die keine Frage von mir je ignorierten. Insbesondere danke ich Inge Achilles, die in Fragen Stipendium immer im Bilde stand. Besonderen Dank auch nochmal an Daniel und Frank, die mir mit ihrem Know-how öfter auch mal privat aushelfen konnten.

Vielen Dank an die IMPRS-Koordinatoren Claudia und (ehemals) Karin, die immer tolle und spannende Kurse im Angebot hatten und sich hervorragend um die Organisation „unseres“ Symposiums kümmerten.

Zu guter Letzt möchte ich meiner Familie danken, die während meiner Doktorandenzeit um das Doppelte wuchs: d.h. meinen beiden Mäusen Jonathan und Elisa und meiner Frau, die ein Leuchtfeuer am Horizont waren und ohne die jegliche Arbeit sinnlos erschienen wäre. Natürlich sei hier auch der Rest meiner Familie erwähnt, allen voran meine Mutter auf die immer Verlass war wenn’s mal wieder eng wurde.

9. Publications and Presentations

Publications

McKenna, DD, Scully, ED, Pauchet, Y, Hoover, K, Kirsch, R, Geib, SM, Mitchell, RF, Waterhouse, RM, Ahn, SJ, Arsala, D, Benoit, JB, Blackmon, H, Bledsoe, T, Bowsher, JH, Busch, A, Calla, B, Chao, H, Childers, AK, Childers, C, Clarke, DJ, Cohen, L, Demuth, JP, Dinh, H, Doddapaneni, H, Dolan, A, Duan, JJ, Dugan, S, Friedrich, M, Glastad, KM, Goodisman, MA, Haddad, S, Han, Y, Hughes, DS, Ioannidis, P, Johnston, JS, Jones, JW, Kuhn, LA, Lance, DR, Lee, CY, Lee, SL, Lin, H, Lynch, JA, Moczek, AP, Murali, SC, Muzny, DM, Nelson, DR, Palli, SR, Panfilio, KA, Pers, D, Poelchau, MF, Quan, H, Qu, J, Ray, AM, Rinehart, JP, Robertson, HM, Roehrdanz, R, Rosendale, AJ, Shin, S, Silva, C, Torson, AS, Jentzsch, IM, Werren, JH, Worley, KC, Yocum, G, Zdobnov, EM, Gibbs, RA and Richards, S (2016) Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. Genome Biol 17: 227.

Busch A., Kunert G., Heckel D.G. and 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, e0184305

Busch A., Kunert, G., Wielsch, N. and Pauchet, Y. (2018) Cellulose degradation in *Gastrophysa viridula* (Coleoptera: Chrysomelidae): functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity. Insect Mol Biol 27: 633-650.

In Preparation

Busch, A., Danchin, E., Pauchet, Y. (2019) Functional analyses of the horizontally acquired Phytophaga glycoside hydrolase family 45 (GH45) proteins reveal distinct functional characteristics

Presentations

Busch A. (2017). Mannanases, but not always! Evolution and functional characterization of glycoside hydrolase family 5 proteins in two species of phytophagous beetles. Talk presented at 16th IMPRS Symposium, International Max Planck Research School, Dornburg, DE

Busch A. (2015). Cellulose-digestion in herbivorous beetles. Talk presented at 14th IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE

Poster Presentations

Busch A., Danchin E.G.J., Pauchet Y. (2018). Cellulases in Phytophaga beetles: Characterization and molecular evolution of family 45 glycoside hydrolases in leaf beetles and weevils. Poster presented at MPI for chemical ecology institute symposium

Busch A. (2017). Cellulases in phytophagous beetles: Characterization and molecular evolution of family 45 glycoside hydrolases in leaf beetles and weevils. Poster presented at 16th Symposium on Insect-Plant Interactions, Tours, FR

Busch A., Heckel D.G., Pauchet Y. (2016). Putative cellulolytic enzymes in *Gastrophysa viridula* and their biological relevance. Poster presented at 15th IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE

Pauchet Y., Kirsch R., Busch A., Vogel H., Heckel D.G. (2014). Evolutionary history of plant cell wall degrading enzymes in phytophagous beetles. Poster presented at Seventh International Symposium on Molecular Insect Science, Amsterdam, NL

Pauchet Y., Busch A., Wurlitzer B., Kirsch R., Heckel D.G. (2014). Digestive adaptations to feed on plants: Origin and function of cell wall degrading enzymes in herbivorous beetles. Poster presented at SAB Meeting 2014, MPI for Chemical Ecology, Jena, DE

Busch A., Heckel D.G., Kirsch R., Pauchet Y. (2014). Cellulose-Digestion in Leaf Beetles. Poster presented at 13th IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE

10. Selbstständigkeitserklärung

Entsprechend der geltenden, mir bekannten Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena erkläre ich, dass ich die vorliegende Dissertation eigenständig angefertigt und alle von mir benutzten Hilfsmittel und Quellen angegeben habe. Personen, die an der Gewinnung der Daten, sowie der Auswertung und der Erstellung der Manuskripte hilfreich waren, wurden namentlich gekennzeichnet. Es wurde weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte für Arbeiten, welche im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen, geldwerte Leistungen erhalten. Die vorgelegte Dissertation wurde außerdem weder als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung noch als Dissertation an einer anderen Hochschule eingereicht.

André Busch
Jena, den 15.12.2018