1 Ancestral Gene Duplication Enabled the Evolution of Multifunctional Cellulases in Stick
2 Insects (Phasmatodea)
3

4 Matan Shelomi^a, David G. Heckel^a, Yannick Pauchet^a
5

6 ^a. Department of Entomology, Max Planck Institute for Chemical Ecology, Hans-Knöll-
7 Str. 8, 07745 Jena, Germany.
8

9 Corresponding Author:

10 Matan Shelomi, Department of Entomology, Max Planck Institute for Chemical Ecology,
11 Hans-Knöll-Str. 8, 07745 Jena, Germany.

12 TEL: +49- 01749614258. EMAIL: mshelomi@ice.mpg.de
13
14

Abstract

The Phasmatodea (stick insects) have multiple, endogenous, highly expressed copies of glycoside hydrolase family 9 (GH9) genes. The purpose for retaining so many was unknown. We cloned and expressed the enzymes in transfected insect cell lines, and tested the individual proteins against different plant cell wall component poly- and oligosaccharides. Nearly all isolated enzymes were active against carboxymethylcellulose, however most could also degrade glucomannan, and some also either xylan or xyloglucan. The latter two enzyme groups were each monophyletic, suggesting the evolution of these novel substrate specificities in an early ancestor of the order. These are respectively the first, true, cellulase-xylanases in the GH9 family and the first xyloglucanases of any kind, let alone bifunctional cellulase-xyloglucanases, to be discovered in Metazoa. Phasmatodea gut extracts could degrade multiple plant cell wall components fully into sugar monomers, suggesting that enzymatic breakdown of plant cell walls by the entire Phasmatodea digestome may contribute to the Phasmatodea nutritional budget. The duplication and neofunctionalization of GH9s in the ancestral Phasmatodea may have enabled them to specialize as folivores and diverge from their omnivorous ancestors. The structural changes enabling these unprecedented activities in the cellulases require further study.

Keywords: Phasmatodea; cellulase; endoglucanase; enzyme; neofunctionalization; glycoside hydrolase

1. Introduction:

Primary plant cell walls comprise cellulose and hemicellulose chains embedded in a pectin matrix (Cosgrove, 2005). Cellulose is a linear chain of β -1,4-linked glucose units and is considered the most abundant organic polymer on earth (Watanabe & Tokuda, 2010). The hemicellulose xyloglucan has a β -1,4-linked glucose backbone with multiple 1,6-xylose residue substitutes, the latter capped with galactose or fucose residues, and is hydrolyzed by distinct enzymes from the cellulases (Vincken et al., 1997; Eklöf et al., 2012). The hemicellulose xylan does not contain glucose, but a backbone of β -1,4-linked xylose [a pentose sugar] residues substituted by a variety of side chains known to include acetyl, arabinosyl, feruloyl, galactose, glucuronosyl, methylglucuronosyl, p-coumaroyl, and rhamnose residues (Pastor et al. 2007). The hemicellulose glucomannan is a polymer of β -1,4-linked mannose and glucose in a 1.6:1 ratio (Katsuraya et al., 2003), and galactomannan has a 1,3- and/or 1,4-linked mannose backbone with 1,6-linked galactose side groups (Ishurd et al. 2004).

To survive on a diet high in solid plant matter, herbivores must utilize plant cell wall degrading enzymes (PCWDEs) to digest some or all of these polymers. Enzymes are either produced by the organism itself (endogenous production) or by a symbiotic microbe associated with the animal. Endogenous and symbiont-produced PCWDEs can work synergistically to fully break down polysaccharides into oligomers and monomers used for energy (Calderón-Cortés et al., 2012), or the enzymes degrade the cell wall just enough to provide access to the easily digestible cytoplasm within. This latter role was thought to predominate in leaf-feeding organisms, whose diets are not as high in cellulose as those of wood feeders.

One well-studied PCWDE family is the glycoside hydrolase family 9 (GH9). Though best known from plants and cellulolytic microbes, GH9s are now thought to have existed in the ancestor to all Metazoan life, having been found in animal genomes from mollusks to tunicates (Davison & Blaxter, 2005). A strong change from traditional views, today scientists increasingly accept that endogenous production of cellulases, wholly independent of gut microbes, is common rather than exceptional, especially in invertebrates (Lo et al., 2011; Fischer et al. 2013). Among insects, most herbivorous orders examined so far have at least one GH9 gene in their genome (Fischer et al., 2013), although as many as four cellulases have been noted in recently sequenced termite genomes (Terrapon et al. 2014). Notable exceptions are the Lepidoptera and Diptera, which lack any cellulases, and many Coleoptera that have cellulases from other GH families (Watanabe & Tokuda, 2011). These include endogenous GH5, GH45, and GH48 found in leaf beetles (Kirsch et al., 2012; Pauchet et al., 2010), which were likely acquired from microbes via successive horizontal gene transfer events (Fischer et al., 2013).

GH9s are almost exclusively endo- β -1,4-glucanases (Enzyme Commission Number (EC) 3.2.1.4) active on carboxymethylcellulose (CMC) and/or amorphous cellulose, though some microbial GH9s with alternative activities on xyloglucan instead of cellulose (endo- β -1,4-xyloglucanase, EC 3.2.1.151) are known (Pauley et al., 1999; Lombard et al. 2014). Bacterial or fungal xylanases (endo- β -1,4-xylanase, EC 3.2.1.8) are generally not known

in the family GH9 (Ahmed et al., 2009; Lombard et al., 2014; Vlasenko et al., 2010), though a bacterial GH9 endo- β -1,4-mannanase (EC 3.2.1.78) capable of degrading mannan and galactomannan is known (Vlasenko et al., 2010). Microbial GH9s that degrade both cellulose and hemicellulosic substrates are rarer, and will often either show relatively low activity on the latter (xylan, in Eckert et al., 2002) or be in multi-enzyme complexes or modular proteins with enzymes from other GH families such as the xyloglucanolytic GH44 (Hirano et al., 2013) or mannanolytic GH5 (Su et al. 2012).

An individual plant GH9 isozyme with concurrent hemicellulolytic functions (active against xylan and glucomannan in addition to CMC) was isolated from rice, but other plants are hypothesized to have similar enzymes (Yoshida & Komae, 2006).

The only record of a biofunctional, cellulolytic and xylanolytic GH9 in animals is a termite GH9 reported to produce large amounts of glucose residues from beechwood xylan when combined with a GH1 β -glucosidase (EC 3.2.1.21) (Scharf et al., 2011), detected using commercial glucose-mutarotase (glucose C2) detection reagents. Including all GH families, the only other cited example (Ni & Tokuda, 2013) of a xylanolytic cellulase in animals, also in a termite but without a GH determination, was claimed based on glucose liberation using a silver nitrate reagent (Potts & Hewitt, 1974). However, in addition to the former involving an enzyme complex, these findings are questionable as xylan does not contain glucose and glucose is not a known product of xylanase activity (Biely, 1985; 2003; Pastor et al. 2007). Positive glucose release from xylan is thus instead conclusive evidence that the substrate is not pure, but is contaminated by a glucose-containing compound (Francisco Javier Pastor, personal communication 2015; Peter Biely, personal communication 2015). Furthermore, even if a xylan product does contain glucose side-chains, these would be generally attached at the 2-O and 3-O positions of the xylose residues and could not, therefore, be removed with an endo- β -1,4-linkage hydrolyzing GH9.

“True xylanases” (Brennan et al., 2004) must act on the xylose backbone to produce pentose sugar poly-, oligo-, or mono-mers. According to a collaborative project organized by the International Energy Agency specifically to identify standardized methods of assaying lignocellulosic enzyme activities, xylanase activity should be measured as detection of reducing xylose, indicating activity on the backbone of the polymer itself (Bailey et al, 1992). This was not done by Potts & Hewitt (1974), and when Scharf et al. (2011) used a bromoaniline detection reagent to detect pentose production, they found no activity, suggesting their enzyme complex is not a true xylanase as defined by any enzymatic authority (Bailey et al., 1992; Brennan et al., 2004; Pastor et al. 2007). Indeed, the authors could find no report on metazoan enzymes claiming to find joint cellulase-xylanase enzymes that also did all of the following: isolate and test individual enzymes, confirm their endogenous origin to the eukaryotic animal or otherwise exclude microbial or endosymbiont enzymes, and use xylose/pentose-based or other xylan-specific assays to appropriately detect xylanase activity. The few studies that did use acceptable methodology, all on termites, did not demonstrate xylanase activity (Ni et al., 2007; Hirano et al., 2010; Zhang et al., 2010). Thus, no conclusive prior evidence exists that any endogenous, animal GH9 can degrade both cellulose and xylan on its own (meaning

not in complex with known xylanases or cellulases), using the universally accepted method of detecting xylanase activity.

In addition, no animal endo- β -1,4-xyloglucanases of any family have ever been reported. Xyloglucanases are known only from plants, bacteria, and fungi, but not Metazoa, although this may be an artefact of how few enzymatic assays in animals have tested for xyloglucanase so far (Benkő et al., 2008; Calderón-Cortés et al., 2012). It should go without saying that bifunctional cellulase-xyloglucanases, xylanase-xyloglucanases, etc, are also unknown in animals. Absence of evidence is not evidence of absence, however, and several unique clades had not yet been analyzed.

Besides leaf beetles, one of the few obligately folivorous insect groups is the Phasmatodea (Calderón-Cortés et al., 2012). Endogenously produced cellulases in this order were isolated and identified as GH9s similar to those of other Polyneopteran insects (Shelomi et al., 2014a). Recent transcriptomic assays of six diverse Phasmatodea species found that each has four to seven distinct and highly expressed copies of the gene (Shelomi et al., 2014b). Low coverage genome sequencing of symbiont-free brain tissue from one of these species produced reads convergent to each of its GH9s. This, combined with all genes' sequence homology to confirmed endogenous GH9s from related insect orders (Polyneoptera) (Shelomi et al., 2014b), the presence of eukaryote-specific signal peptides, the finding of these enzymes in multiple Phasmatodea species raised on different diets and in different countries, and the lack of microbial symbionts or morphological adaptations for microbial fermentation in Phasmatodea digestive systems (Dillon & Dillon, 2004; Shelomi et al., 2013, 2015), all supported the conclusion that the Phasmatodea GH9s so far described are all endogenously produced. This finding is completely in line with contemporary theories on the evolution of animal cellulases (Davison & Blaxter, 2005). Cellulase genes are also present in the genome of the Phasmatodea *Timema cristinae* (Shelomi et al., 2014b; Soria-Carrasco et al., 2014) (http://nosil-lab.group.shef.ac.uk/?page_id=25). To reiterate: the “previous dogma” (Watanabe & Tokuda, 2011) that insects depend on microbes for cellulases has been overturned, and the fact that the Phasmatodea produce cellulases independently of their microbiota is thoroughly and exhaustively supported by multiple lines of evidence, including thorough and exhaustive rejection of the alternate hypothesis of microbe-dependant cellulase production, and is in no way a challenge to any theory of insect cellulases produced in the last decade.

What is unusual, however, is the number of gene copies, at nearly half a dozen per species. Phasmatodea also express numerous copies of β -glucosidase, a ubiquitous animal enzyme that is part of the insect cellulase complex, breaking down cellobiose and/or other sugar dimers into glucose (Calderón-Cortés et al., 2012). Phasmatodea also possess a heavily ridged proventriculus lined with small spinules that is capable of reducing ingested plant matter to pulp (Shelomi et al., 2015). These features suggested that Phasmatodea were a likely candidate for symbiont-independent, complete cellulose degradation into glucose (Calderón-Cortés et al., 2012; Fisher et al., 2013), but this had never been demonstrated. Assaying the activity of extracted gut lumen proteins against cellulosic polymers would test this hypothesis, but would not answer the question of why

so many GH9 endoglucanase genes are expressed in the gut and why they have been maintained through evolutionary time. The goal of our investigations was to express the individual GH9 enzymes identified from the Phasmatodea transcriptomes and test their activities independently against carboxymethylcellulose (CMC) and other plant cell wall components, with the hypothesis that some copies had evolved novel or additional functions to cellulase activity that increase the insects' ability to digest leaves, and which would explain the preservation of multiple GH9s in the Phasmatodea genomes.

2. Materials and Methods:

2.1 Screening for putative PCWDE transcripts

The six species with published transcriptomes (Shelomi et al., 2014b) are *Aretaon asperrimus* (Heteropterygidae), *Extatosoma tiaratum* (Phasmatidae), *Medauroidea extradentata* (Phasmatidae), *Peruphasma schultei* (Pseudophasmatidae), *Ramulus artemis* (Phasmatidae), and *Sipyloidea sipyilus* (Diapheromeridae). We also added *Timema cristinae* (Timematidae), due to its position in Timematodea, the sister group to all other stick insects, the Euphasmatodea (Beutel et al. 2013; Friedemann et al., 2012). Euphasmatodea split from Timematodea between 125 (± 60) and 103 (± 19) million years ago (Misof et al., 2014; Bradler et al., 2015), and Euphasmatodea began to differentiate into its current species approximately 61 (± 14) million years ago (Bradler et al., 2015). Thus the *Timema* data would allow us to study the evolution of these enzymes across the entire order, and enable us to draw conclusions about the enzymes in the common ancestor of these clades.

Amino acid sequences from known insect and microbial cellulases were retrieved from Genbank and used with tBLASTn (e-value cutoff 10^{-10}) to mine for homologous sequences from the published transcriptomes (Shelomi et al., 2014b) (GenBank Accession No. PRJNA238833 & PRJNA221630) and the *Timema cristinae* genome (Soria-Carrasco et al., 2014) (http://nosil-lab.group.shef.ac.uk/?page_id=25, BioProject PRJNA243533). We also BLASTED putative sequences from *P. schultei* against the genomic reads from the *P. schultei* brain (GenBank Accession No. SRX480713) that had been produced to confirm endogeneity of any putative enzymes (Shelomi et al. 2014b).

2.2 Generating cDNAs and cloning of full length genes

We dissected midguts, removing their contents to reduce plant and microbe contamination, from cultured specimens of the six Euphasmatodean species. We also analyzed wild-caught *Timema cristinae* (Paradise Reserve, Santa Barbara, CA). The tissue was stored in RNAlater[®] solution (Qiagen, Venlo, Netherlands). From midgut tissue macerated in a frozen tissue Lyser, RNA was extracted using the innuPREP RNA MiniKit (Analytik-Jena, Jena, Germany) and purified with theRNeasy[®] MinElute[®] cleaning kit (Qiagen) as per manufacturers' instructions. From the RNA, cDNA was extracted and RACE-PCR performed as needed with the SMARTer RACE cDNA Amplification Kit (BD Clontech, Saint-Germain-en-Laye, France) as per manufacturer's instructions. If transcripts were incomplete, specific primers were produced for 5'- and 3'-Rapid Amplification of cDNA Ends (RACE) PCR using the Primer3 program v0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>). PCR products were cloned into One Shot[®] Top10 Chemically Competent *E. coli* cells with the pCR[™]4-TOPO/TA[®] Vector

(Invitrogen, Carlsbad, CA), and sequenced. Once complete open reading frames (ORFs) were available for every endoglucanase gene, they were annotated and submitted to Genbank (Accession Numbers KT338252-KT338292, Table S1).

2.3 Phylogenetic analysis

ORFs were converted to amino acid sequences and checked for eukaryote-specific signal peptides using the SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>), which were then removed. For both the Phasmatodea and Metazoa GH9 trees, amino acid alignments using MUSCLE v3.8.31 (Edgar, 2014) were corrected manually when needed and used to estimate phylogenetic relationships using maximum likelihood analysis in MEGA6 (Tamura et al., 2013). The best protein evolution model was determined using the built-in “find best DNA/protein models” tool, and identified as ‘Whelan and Goldman’ incorporating a discrete gamma distribution and a proportion of invariable sites (WAG+G+I). Each analysis was tested for robustness using 1000 bootstrap replicates. Conserved GH9 domains were identified in the aligned sequences manually. To produce posterior branch probabilities, Bayesian analysis was performed with MrBayes v3.1.2 (Ronquist et al., 2012), generating one million realizations using a burn-in rate of 25% and sampling every 100 generations.

2.4 Expression in insect cells

M. extradentata, *P. schultei*, *S. sipylus*, and *T. cristinae* were chosen for individual enzyme analyses (see results). For these, gene specific forward and reverse primers (Table S1) were designed to amplify the complete ORF of each putative enzyme from the cDNA and the resulting PCR products were cloned into the pIB/V5-His TOPO/TA[®] vector (Invitrogen) in frame with the C-terminal V5-His tag. Kozak sequences were included at the beginning of the forward primers, and the stop codon removed from the end of the reverse primers. Colony PCR or direct sequencing was done to ensure the genes were transformed in the correct direction, and the plasmids extracted with a GeneJET[™] Plasmid Miniprep Kit (Thermo Scientific, Waltham, MA) and transfected into Sf9 cells (Invitrogen) using the reagent FuGENE HD (Promega, Madison, WI) under a laminar flow biosafety hood. Negative controls of Sf9 cells without plasmid were prepared as well. Culture medium was harvested after 72 hours and centrifuged, and the supernatant tested for successful expression via Western blot (Figure S1) with anti-V5-HRP antibody (Invitrogen).

2.5 Plate assays for substrate activity

To test the activity of the expressed enzymes on different cellulose and hemicellulose substrates, square Petri dishes were made with 0.1% solutions of substrate in 0.4% agarose and 50 mM citrate-phosphate buffer (pH 5.0). Substrates were CMC (Sigma-Aldrich, St. Louis, MO), xylan from beechwood (Sigma-Aldrich), xyloglucan from tamarind seed (Megazyme, Wicklow, Ireland), glucomannan from konjac (Megazyme), and galactomannan from carob (Megazyme). The pH was chosen based on the average pH of the Phasmatodea anterior midgut where cellulases are most highly expressed (Shelomi et al., 2014b, 2015), however a pH series was made with CMC to test the ideal pH for each enzyme (Figure S2). Wells were made in the plates and filled with 5µL of protein (or 10µL for the xylan plates), incubated upside-down at 40°C overnight, stained

for one hour in 0.1% Congo Red (Colour Index No. 22120) in 0.1 M Tris-HCl buffer pH 8.0, and destained in 1 M NaCl. Enzyme activity was detectable as clearings in the stained gel. Sterile technique and negative controls were used to exclude the possibility of contamination of the wells.

2.6 Thin Layer Chromatography (TLC) assays for substrate activity

To complement the results of the plate assays, enzyme solutions were dialyzed in three baths of 50mM citrate-phosphate buffer pH 5.0 at 4°C using Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific) with 10KDa cutoffs, desalted in Zeba Desalt Spin Columns (Thermo Scientific) with a 7KDa cutoff, and stored at 4°C until use. Ten microliters of desalted enzyme were combined in microcentrifuge tubes with 1% w/v substrate in water in the volumes as in Table S2, 2μL 0.2M citrate phosphate buffer, pH 5.0, and enough diH₂O to produce 20μL total. To test for activity on crystalline cellulose we used Avicel (Sigma-Aldrich) and filter paper (Whatman, Maidstone, UK). For the latter, a 1mm² piece was soaked in buffer, then moved to the buffer+[enzyme] solution. The tubes were incubated for 16hrs at 40°C, then spotted onto TLC plates (silica gel 60, 20x10 cm, Merck KgaA, Darmstadt, Germany) and developed with 2:1:1 of 1-butanol/glacial acetic acid/water. Reference standards for the cellulosic substrates and xyloglucan were 2μg each of glucose, cellobiose, cellotriose, cellotetraose, and cellopentaose (Sigma-Aldrich). For xylan and xyloglucan, reference standards were 2μg each of xylose (Sigma-Aldrich), xylobiose, and xylotriose (Megazyme). For mannans, the reference standards included the glucose standards and 2μg each of D-mannose, mannobiose, mannotriose, mannotetraose, mannopentaose, and mannohexaose (Sigma-Aldrich). A reference spot of 2μg of galactose (Sigma-Aldrich) was added for galactomannan. The positive control was cellulase from *Trichoderma reesei* ATCC 26921 (Sigma-Aldrich). The dried plates were sprayed with 0.2% (w/v) orcinol in 9:1 methanol/sulfuric acid, then warmed with a heat gun until spots appeared. Images of the TLC plates are Figures S3, S4, S5.

2.7 Whole gut digestome testing

To test what all the enzymes in the insect gut can do together, including endogenous Phasmatodea enzymes other than the GH9s as well as any ingested plant or microbial enzymes, whole anterior midguts minus contents of up to five adults for each species were collected from live, cultured specimens and macerated in a 100mM citrate phosphate buffer, pH 5.0 with dissolved cOmplete, Mini, EDTA-Free protease inhibitor cocktail (Roche Diagnostics, Basel, Switzerland). After centrifugation, the supernatant was collected and applied to plate assays as above. The results of this assay would show what the net activity of the total Phasmatodea digestome produces from each substrate, with the production of monomers indicating complete polymer breakdown into assimilable sugars and suggesting a nutritive role for the plant cell walls. In addition to positive and negative controls, macerated blackberry leaves were also tested to control for plant-derived enzymes in the diet (Figure 1). To specifically look at enzymes secreted into the exoperitrophic space, thus excluding enzymes contained within the midgut cells themselves, three male *P. schultei* were dissected, their food boluses contained within the peritrophic membranes removed, and the anterior midgut soaked in 600μL citrate phosphate with protease inhibitor. This mixture was centrifuged and the supernatant dialyzed, desalted, and used for TLC analysis as above (Figure 2).

3. Results:

3.1 Total digestome and midgut lumen enzyme activity

The macerated midguts of the six original species showed activity by agarose diffusion assays on CMC, xylan, xyloglucan, glucomannan, and galactomannan far exceeding residual enzymes from their plant diet (Figure 1). These midgut samples contained any enzymes from the gut microbiota and ingested leaves, in addition to all enzymes produced endogenously by the insect itself and those which are not secreted by midgut tissue, and so represent the total enzymatic complement of each species' digestome (Fischer et al. 2013). The endogenous plant enzymes or plant surface microbe enzymes alone (BB in Figure 1) could not degrade xylan or galactomannan at all, and were considerably weaker on xyloglucan and glucomannan than the insect gut samples or positive control. This confirms that the Phasmatodea midgut contains active enzymes that are produced by the insect itself and/or by gut microbes, and which are capable of degrading the five above plant cell wall polymers.

Midgut lumen fluid devoid of gut content taken from *P. schultei*—the only species with quantified enzyme expression levels (Shelomi et al., 2014b)—was tested against multiple cellulosic and hemicellulosic polysaccharides via thin layer chromatography (TLC) (Figure 2). These *P. schultei* lumen samples should be devoid of diet-derived enzymes that cannot pass the peritrophic membrane, and devoid of midgut tissue enzymes that are not secreted. The samples were detectably active on every substrate tested except Avicel, a crystalline form of cellulose, and galactomannan. Glucosic compounds from cellulose (CMC) to cellobiose (Glu2) could all be degraded into the monomer, confirming the existence of endo- β -1,4-glucanases and β -glucosidases, both of which are found in the Phasmatodea midgut transcriptomes (Shelomi et al. 2014b). Some glucose dimers and monomers were produced from the filter paper, suggesting the existence of an exo-cellulase (exo- β -1,4-glucanase, EC 3.2.1.91), which cleaves dimers from the ends of cellulose chains. However, no such enzyme was found in the transcriptomes or genomes.

In addition, xylan (Xyl), xylotriose (Xyl3), and xylobiose (Xyl2) could all be degraded into the xylose monomer (Figure 2), suggesting the presence in the midgut of true endo- β -1,4-xylanase to break down the polymers and β -1,4-xylosidase (xylobiase, EC 3.2.1.37) to break down dimers and trimers. However, no typical xylanases or β -xylosidases were found in the *P. schultei* or other transcriptomes or *Timema* genome. Likewise, xyloglucan (XyloG) was broken into large oligomers and xylose monomers (Figure 2), suggesting the presence of endo- β -1,4-xyloglucanases or xyloglucan-specific endo- β -1,4-glucanases to break down the polymer and α -1,6-xylosidases (EC 3.2.1.177) to remove the xylose side chains. The latter were found to be endogenously produced in Euphasmatodea, but these are common metazoan enzymes so their presence is not surprising. Glucomannan (GluM), mannohexaose (Man6), and mannobiose (Man2) were degraded to mannose monomers (Figure 2). The polymer can be degraded to oligomers by endoglucanases or endomannanases, the latter not found in the Phasmatodea transcriptomes or genomes. The oligomers and dimers can be broken by β -mannosidase (E.C. 3.2.1.25) and β -1,4-mannobiosidase (EC 3.2.1.100), which are found in Phasmatodea as well as throughout

Insecta. The absence of activity on galactomannan (GalM) (Figure 2) suggests an absence of true mannanases, as galactomannan has a purely mannose backbone.

Thus the total gut activity suggests greater PCWDE diversity, namely large polymer hydrolases, than the transcriptomes initially suggest. Since only genes encoding putative endo- β -1,4-glucanases (GH9 cellulases) but no genes encoding hemicellulolytic enzymes could be identified from Phasmatodean transcriptomes or *Timema* genome, and since the diet does not contribute xyloglucanases, xylanases, or mannanases significantly or at all (Figure 1), we then asked whether some of the GH9 enzymes were responsible for breaking down hemicellulose polymers, either in addition to or instead of degrading cellulose.

3.2 Phylogenetic analysis of endogenous GH9s

Complete open reading frames were successfully obtained for all the endogenous, GH9 cellulases from the six transcriptomes, plus one new *S. sipylus* cellulase gene that the RNAseq approach could not differentiate from a similar transcript (Table S1). We also searched for GH9 cellulases from the draft genome of *Timema cristinae* (Soria-Carrasco et al., 2014). We identified five (Table S1), four of them located in tandem on scaffold 24, which has been mapped to linkage group 4 of the *Timema* genome, and one (TCR4) located on the unmapped scaffold 551. These five *T. cristinae* GH9 genes possess a similar structure comprised of nine exons (Soria-Carrasco et al., 2014). The products of all Phasmatodea GH9 genes are all similar in size (449 to 453 amino acids, except TCR4 at 467), meaning all are single domain enzymes rather than multi-enzyme proteins. All are closely related to other insect GH9 cellulases as revealed by BLAST analyses, with similar amino acid lengths (Watanabe & Ni, 2013) and a phylogenetic analysis (Figure 3) shows that Phasmatodea GH9s form a monophyletic clade sister to other insects and hexapods GH9s, further confirming their already well established endogenous origin and evolution from ancestral insect genes (Davison & Blaxter, 2005). The amino acid sequences of every protein contained the conserved, catalytic domains of GH9 enzymes (Khademi et al., 2002), further confirming their identity as GH9 enzymes (Zhou et al., 2004; Davison & Blaxter, 2005). Based on these domains alone, all enzymes are expected to have similar, endo- β -1,4-glucanase activity. Maximum-likelihood phylogenetic analysis of the aligned Phasmatodea sequences (see Methods) revealed four highly supported branches (Figure 3,4). Each Euphasmatodea species has at least one gene in each group, with *Timema* genes appearing as the sister genes to each cluster of euphasmatodean GH9s.

3.3 Expression and substrate specificity of individual GH9 enzymes

From the Euphasmatodea, *M. extradentata*, *P. schultei*, and *S. sipylus* were chosen for individual enzyme analyses based on the phylogenetic results. All their cellulase enzymes were successfully cloned and heterologously expressed in insect Sf9 cells (Figure S1). Their activities are summarized in Figure 4 (See also Figures S3-5 for TLC assays and S5-6 for plate assays). The ideal pH for enzyme activity was around 5.0 (Figure S2), which, as hypothesized, matches that of the anterior midgut where the genes encoding these enzymes are most highly expressed (Shelomi et al., 2015).

All Euphasmatodea GH9s tested showed positive activity on CMC, and most showed some activity on glucomannan, meaning all are endo- β -1,4-glucanases (Figure 4). Some enzymes from each species could break down glucose oligomers, with a few (PSC1, MEC3) showing some activity on cellotriose, however no enzyme could degrade glucose dimers. Activity on filter paper was weak, producing glucose dimers and trimers, with nearly imperceptible activity on Avicel. Such ability is known in GH9s from insects, namely termites, though like the phasmid enzymes they are weakly active on these substrates (Ni et al., 2005). Enzymes with the inability to hydrolyze filter paper, cellopentamers, and/or cellotetramers clustered together phylogenetically (Figure 4).

In addition to their activity against cellulose, some individual GH9 enzymes were capable of degrading hemicelluloses (Figure 4), producing clearings on the agar substrate plate assays (Figures S6-7) and bands on the TLC assays (Figures S3-5). Four enzymes (MEX3, PSC3, SSI3, SSI7) could degrade xylan down to oligomers as small as xylotriose. These enzymes formed a monophyletic group along with representatives from the other Euphasmatodea. For xyloglucan, eight enzymes (MEX2, MEX4, MEX5, PSC4, PSC5, SSI5, SSI6, SSI8) could break the polymer down to large oligomers. These enzymes too formed a well-supported monophyletic group with three subgroups. Multiple enzymes could also degrade glucomannan, producing oligomers no smaller than trimers. Enzymes with strong, weak, or no glucomannolytic activity all clustered phylogenetically. Because no enzyme could degrade galactomannan or mannohexaose, we cannot claim that the GH9s degrading glucomannan are true endo- β -1,4-mannanases, for instead they may be exclusively acting on the β -1,4-glycosidic bonds of the glucose residues also present in the glucomannan backbone.

From *T. cristinae*, apart from TCR5, all enzymes were successfully cloned and heterologously expressed in *Sf9* cells (Figure S7), and their results are summarized in Figure 4 (See Figure S7 for plate assays). All enzymes showed activity on at least two plant cell wall polysaccharides, except TCR4, which showed no activity against any of the substrates tested. Put into a phylogenetic context, the *Timema* GH9 active on xyloglucan (TCR1) clustered as a strongly supported sister branch to the Euphasmatodea-derived enzymes also active on xyloglucan. Similarly, TCR2, active on xylan, clustered as a strongly supported sister branch to the Euphasmatodea-derived enzymes also active on xylan. Finally, TCR3 clustered as a strongly supported sister group to the Euphasmatodea-derived enzymes active on glucomannan but not on the other hemicellulases (Figure 4). Altogether, our results indicate that the diversification of the GH9 enzymes in Phasmatodea and the ability of some to break down hemicellulolytic substrates along with amorphous cellulose evolved in or before the last common ancestor of the Timematodea and the Euphasmatodea.

4. Discussion:

4.1 Novel, Bifunctional and Multifunctional Enzymes

The most significant result of our work is the discovery of unprecedented substrate specificities and biofunctionalities. These Phasmatodea enzymes represent the first documented case in Metazoa of GH9 endo- β -1,4-glucanases individually [not in complex and not as part of large, multi-domain proteins] capable of breaking down xylan,

xyloglucan, and/or glucomannan— using the accepted definition of true polymer hydrolase activity, in which the polymer backbone of β -linked units is degraded into its component oligomers— all while maintaining their cellulolytic ability. Even the latter was unusual: weak activity on cellotriose was noted for most of the individual GH9s (especially MEX3 and PSC1), with some exceptions most insect cellulases normally cannot hydrolyze oligosaccharides smaller than cellotetraose (Watanabe & Ni, 2013).

We have successfully demonstrated true endo- β -1,4-xylanase activity in five Phasmatodean endo- β -1,4-glucanases, with three enzymes (AAS3, ETI3, RAR2) putatively sharing this bifunctionality based on their phylogenetic relatedness (Figure 4). These endogenous, bifunctional, cellulose-xylan degrading enzymes (EC 3.2.1.4 & 3.2.1.8) are a first in Metazoa. Our enzymes are also likely the first bifunctional endoglucanases/endoxylanases among all non-plant GH9s (including microbial GH9s), although cellulose-xylan degrading enzymes from other GH families have been possibly detected in insect microbial symbionts (GH5 in Scrivener et al., 1997; GH7 in Zhou et al., 2007) and bacteria (GH44 in Kim et al. 2008, Warner et al., 2011).

We have successfully demonstrated true endo-1,4-xyloglucanase activity in nine Phasmatodean endo- β -1,4-glucanases, with six putatively bifunctional enzymes based on phylogenetic relatedness (Figure 4). These endogenous xyloglucanases are unambiguously the first to be described in any animal, representing a remarkable development for Metazoan enzymology. That the first animal xyloglucanases are also bifunctional cellulases is also greatly unexpected. While it is not surprising that an endo- β -1,4-glucanase can degrade xyloglucan, as in xyloglucan-specific endo- β -1,4-glucanases (EC 3.2.1.151), what is surprising is the retention of activity against cellulose (EC 3.2.1.4). We could not find any record in the literature of an individual enzyme with only a single, GH9 catalytic domain that has bifunctionality against cellulose and xyloglucan, including the rice GH9 (Yoshida & Komae, 2006), however at least one multifunctional, bacterial GH44 is known (Warner et al., 2011). We have thus confirmed the hypothesis of Eklöf et al. (2012) that such endo- β -1,4-xyloglucanase activity among endo- β -1,4-glucanases would be found in emerging enzyme discovery projects, although even that paper did not mention animals as a possible source.

We have successfully demonstrated true polymer hydrolysis activity on glucomannan in several enzymes, with the strongest activity clustered phylogenetically in two groups. However, we cannot determine whether this breakdown is in the β -1,4 bonds between two glucoses, between two mannoses, or between glucose and mannose in the glucomannan backbone. Due to the inability of any enzyme to degrade mannohexaose or galactomannan, we cannot claim to have found true mannan endo- β -1,4-mannanases. Rather, we believe our endo- β -1,4-glucanases are bifunctional against cellulose and glucomannan [and sometimes also xylan, as in MEX5 and SSI5, or xyloglucan as in MEX3 and PSC3]. This, too, is unknown in Metazoa, but documented in GH9s from rice (Yoshida & Komae, 2006) and bacteria (Vlasenko et al., 2010).

A previous paper argued that, rather than attack the polymer backbone itself, GH9s can remove side chains from polysaccharides possessing them (Scharf et al., 2011). One

could thus argue that our multifunctional enzymes are standard cellulases that happen to remove sugars from hemicelluloses coincidentally, but that is incorrect. Not only do we reject that paper's definition of hemicellulolytic activity as including side-chain removal (Bailey et al., 1992; Brennan et al., 2004; Pastor et al., 2007), but also we have not observed such activity here. First, if only side chains were being removed from hemicellulolytic substrates, then the agar plate assays of the isolated enzymes would not have revealed clearings confirming the breakdown of the polymer itself (Figures S6-7). Second, if any glucose residues from xylan, xylose residues from xyloglucan, or any of the variety of other monomer residues on either (Eklöf et al., 2012; Katsuraya et al., 2003; Pastor et al., 2007) were being removed at all, then we would have seen them on the TLCs, which we did not (Figures S3-5). Instead we only saw oligomers resulting from the breakdown of the polymers' backbones. For xylan in particular, we saw xylose oligomers, which can be further distinguished from glucose oligomers by their blue color on our TLCs relative to glucosic molecules' purple (Figure 2). Third, side chains are usually attached via α -1,6-linkages, so it is unlikely for an $\text{endo-}\beta$ -1,4-linkage hydrolyzing enzyme such as a cellulase to be able to remove side chains at all (although, given our own results overturning conventional enzymology wisdom, we cannot say for sure that such a bifunctional endoglucanase is impossible). Lastly, that the xylanolytic and xyloglucanolytic enzymes each form strongly-supported monophyletic groups suggests selected evolution for that ability in each group (Figure 4). Certain Phasmatodea GH9s are thus independently capable of true hydrolysis of both hemicellulose and cellulose, and nothing like this has been reported before outside of rice (Yoshida & Komae, 2006).

Of course, the possibility exists that other animals, including non-insects, also have xyloglucanases and even bifunctional cellulase-xyloglucanases and cellulase-xylanases. Either these organisms have not yet had their enzymes analyzed, or they have, but not against both cellulosic and hemicellulosic substrates. This discovery in the Phasmatodea demonstrates the importance of thorough enzymatic analysis [with the proper methodology] of a larger variety of organisms. In particular, the assumption that only plants and microbes produce xyloglucanases, like the equivalent belief that endogenous insect cellulases do not exist, is now disproven. This finding means xyloglucans should be added to future qualitative assays of animal enzyme substrate specificities, as others have recommended for microbes (Benkö et al., 2008; Eklöf et al., 2012).

4.2 Evolution of Phasmatodea Endoglucanases

The Phasmatodea GH9s cluster phylogenetically into four paralogous groups inherited from an ancestor preceding the Timematodea/Euphasmatodea split, one group of which is xyloglucanolytic and another xylanolytic. Some further gene duplication events occurred after the Timematodea and Euphasmatodea split, but before the six Euphasmatodea species diverged. In total, seven strongly supported clusters can be found, orthologs of which show similar [in]abilities to degrade CMC, xylan, xyloglucan, and glucomannan. We predict that enzymes from the three untested species will have similar enzymatic capabilities as their orthologs, and that enzymes from other Phasmatodea not tested here will also fall into these seven clusters and show similar digestive abilities. The possible presence of these GH9 enzyme groups, with or without hemicellulolytic ability, in any of

the putative sister groups to the Phasmatodea—the Embioptera and Xenonomia, neither of whose digestive enzymes have been studied—would not only help time the gene duplication and neofunctionalization events and correlate them to other physiological traits, but also resolve the still debated relationships of the Polyneoptera phylogenetic tree (Beutel et al., 2013). That the known termite GH9s and Phasmatodea GH9s formed separate monophyletic groups (Figure 3) suggests that the four paralogous Phasmatodea groups were not inherited from a common ancestor of the Phasmatodea, Embioptera, and Xenonomia with the Dictyoptera (the monophyletic clade containing termites, cockroaches, and mantids) (Beutel et al., 2013), so from this tree one cannot predict whether or not Dictyopteran GH9s would also have hemicellulolytic abilities.

The stable preservation of many copies of the same gene family is thought to require divergent functions. Following gene duplication, multiple functions of the original gene may be split among the duplicates (“subfunctionalization”), or a duplicate may develop a completely new function (“neofunctionalization”). Note that neofunctionalization (NF) traditionally implies that the new function replaces the original, but we are using He & Zhang’s (2005) definition in which the copies may retain all (NF-I), none (NF-II) or some (NF-III) or the original gene’s functions. Within each Phasmatodea species, most of these gene copies are highly expressed (Shelomi et al., 2014b), which makes better sense if each has a different function. As hypothesized, some of the Phasmatodea GH9s have developed the novel ability to degrade hemicellulose polysaccharides while retaining most or all of the ancestral ability to degrade amorphous cellulose: a type I neofunctionalization (He & Zhang, 2005). According to the functional characterization of the *T. cristinae* GH9s, xyloglucanolytic and xylanolytic functions each evolved once in an ancestor of the Timematodea and Euphasmatodea, meaning an ancestor of the entire order Phasmatodea [and possibly other Polyneoptera].

Could the multifunctional abilities of the Phasmatodea enzymes be retained or possibly recovered functions of an even older enzyme, possibly predating the evolution of insects? This is highly unlikely given the lack of such abilities in the known metazoan GH9s, all of which share a common ancestor (Davison & Blaxter, 2005), nor would this hypothesis fit parsimoniously with the evolution of the insect GH9s (Figure 3) and Phasmatodea GH9s alone (Figure 4). All evidence suggests the Phasmatodea GH9s are monophyletic, as are the xylanolytic and xyloglucanolytic groups within. Again, we are not claiming that these genes evolved in the common ancestor of the Phasmatodea *sensu stricto*, as the Embioptera and Xenonomia enzymes must be studied before making any such conclusions. The possibility that these hemicellulolytic GH9s exist deeper into the Polyneoptera phylogeny, including in the termites, cannot be overruled due to the lack of proper tests for xylanase or xyloglucanase activity in these insects. The glucomannanolytic abilities may be the most ancestral, and testing other insect GH9s for such abilities is necessary.

4.3 Implications of the Phasmatodea Digestome

Besides the GH9s, multiple other insect-, microbial, and food plant-derived enzymes should be active in the Phasmatodea digestive tract, although the need for symbiont-derived hemicellulases can now be ruled out. In particular, the production of sugar

monomers from the various oligomers produced by the GH9s, as detected by the midgut lumen TLC (Figure 2), demonstrates the existence of synergism between multiple PCWDEs leading to complete breakdown of plant cell walls into nutritive sugars. The degradation of cellobiose into glucose monomers has been documented in Phasmatodea before and is almost certainly performed by their endogenous β -glucosidases (Shelomi et al., 2014b). Alternatively, production of glucose from microcrystalline cellulose via transglycosylation has been noted and may be responsible for the glucose monomers in the digestome sample (Klesov, 1990).

It remains unknown what enzyme in the *P. schultei* lumen is degrading the larger mannose oligomers, xylotriase, or xylobiose (Figure 2), or what enzyme[s] in the midgut tissue (Figure 1) but not the lumen can degrade galactomannan. Since the lumen sample would also contain any enzymes produced by transient microbes in the gut (Shelomi et al., 2013) and the gut extract contains proteins that do not leave the tissue cells, the mannanase activities may be due to non-secreted proteins. All of these functions could be attributed to microbial enzymes, or to yet more undescribed neofunctionalizations in other Phasmatodea genes.

This work also provides compelling evidence that insects, or at least Phasmatodea, are an equally or more promising source for novel PCWDEs with diverse capabilities for industrial usage, especially in the production of biofuels (Oppert et al., 2010; Watanabe & Ni, 2013), which has focused mostly on microbes (Morrison et al., 2009). Bifunctional enzymes in particular are highly-sought after in this regard due to their ability to break down a larger proportion of the plant feedstock (Benkő et al., 2008; Su et al., 2012). That our bifunctional enzymes work without complexes and are small proteins compared to the multidomain enzymes is an advantage for possible expression of these enzymes in a microbial vector better suited for an industrial biofuel reactor. Understanding how the structure of these enzymes enables their expanded substrate specificity would also be of interest. However, we were unable to link the novel substrate specificities to any amino acid substitutions at catalytically important residues (Khademi et al., 2002), as had been posited for variant cellulases in sea squirts (Davison & Blaxter, 2005). X-ray crystallography analysis on some of these enzymes, in particular MEX3 from the xylanases and MEX5 from the xyloglucanases, is the next logical step in determining how these unusual activities are possible.

We hypothesize that the duplication and neofunctionalization of Phasmatodea GH9s may have greatly impacted the evolution of the order, correlating with or causing the shift to obligate folivory. The ability to degrade multiple plant cell wall components would greatly facilitate access to the nutrient- and nitrogen-rich cytoplasm within the cells of the leaves they digest. We also now have evidence that Phasmatodea can fully degrade plant cell wall components into monomers that could be used as energy, and their large number of endogenous β -glucosidases strongly implies that their digestion of cellulose to glucose is entirely microbe-independent. Complete folivory without occasionally cheating on one's diet and engaging in carnivory or necrophagy, as documented frequently in grasshoppers (Richardson et al., 2012), likely could not have evolved in Phasmatodea without the neofunctionalizations from their ancestral cellulases. Experimental

manipulation of some or all of these Phasmatodea genes and subsequent observation of their feeding rate, survival, and possible acceptance of non-leaf diets may help test the importance of enzyme diversity on maintaining such a specialist diet.

An analogous evolutionary event is the acquisition by horizontal gene transfer of multiple GH family enzymes (but never GH9s) in the leaf beetles (Chrysomelidae) (Eyun et al., 2014; Kirsch et al., 2012), which had lost their GH9s earlier in Coleoptera evolution. Chrysomelidae and Phasmatodea are the only major clades of insects that are obligately folivorous in every life stage (Calderón-Cortés et al., 2012), and both also have an unusually large complement of PCWDEs that can digest multiple types of substrate. The difference is that the beetles acquired multiple enzyme family genes from external sources, while the Phasmatodea evolved new functions in their one cellulase family on their own. These two paths to greater enzyme diversity may have enabled these clades to specialize entirely on leaves. In phasmids, this would explain the origin of fully herbivorous groups whose closest relatives are detritivores and carnivores (Misof et al., 2014). Though presently speculative, with this “Enzyme Expansion Hypothesis” for the evolution and maintenance of dietary specialization such as folivory, we hope to explain the evolutionary divergence of groups like Phasmatodea from a mechanistic perspective.

Acknowledgements:

Preserved Euphasmatodea tissue was provided by Lynn Kimsey and Steve Heydon of the Bohart Museum of Entomology, University of California-Davis (USA), and *Timema cristinae* by Cristina Sandoval of Paradise Reserve, University of California-Santa Barbara (USA). Live Phasmatodea came from Benyo Andras of the Budapest Zoo (Hungary) and from private collectors Sacha Eilmus, Corinna Krempel, and Manuel Wagner (Germany). Special thanks to the Phasma Culture list for facilitating Phasmatodea acquisition. Thanks to staff at the Max Planck Institute for Chemical Ecology (Jena, Germany), in particular André Busch, Roy Kirsch, Henriette Ringys-Beckstein, Domenica Schnabelrauch, and Bianca Wurlitzer. This research was funded by the National Science Foundation (USA) Postdoctoral Research Fellowship in Biology, Grant No. DBI-1402883, funds from the University of California, Davis (USA) and by the Max Planck Society (Germany).

Competing Interests:

The authors have no financial or non-financial competing interests.

References:

- Ahmed, S., Riaz, S., Jamil, A., 2009. Molecular cloning of fungal xylanases: an overview. *Appl. Microbiol. Biotechnol.* 84(1), 19-35. doi:10.1007/s00253-009-2079-4
- Bailey, M. J., Biely, P., Poutanen, K., 1992. Interlaboratory testing of methods for assay of xylanase activity. *Journal of biotechnology.* 23(3), 257-270.
- Benkő, Z., Siika-aho, M., Viikari, L., and Réczey, K., 2008. Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates. *Enzyme Microb Technol.* 43(2), 109-114.
- Beutel, R.G., Wipfler, B., Gottardo, M., Dallai, R., 2013. Polyneoptera or "Lower Neoptera" - New Light on Old and Difficult Phylogenetic Problems. *Atti Accademia Nazionale Italiana di Entomologia.* 61, 113-42.
- Biely, P., 1985. Microbial xylanolytic systems. *Trends in Biotechnology.* 3(11), 286-290.
- Biely, P., 2003. Xylanolytic enzymes. *Handbook of Food Enzymology.* New York, Marcel Dekker, Inc.: 879-915.
- Bradler, S., Cliquennois, N., Buckley, T.R., 2015. Single origin of the Mascarene stick insects: ancient radiation on sunken islands? *BMC Evolutionary Biology.* 15(1), 196.
- Brennan, Y., Callen, W.N., Christoffersen, L., et al., 2004. Unusual microbial xylanases from insect guts. *Appl Environ Microbiol.* 70(6), 3609-3617.
- Calderón-Cortés, N., Quesada, M., Watanabe, H., Cano-Camacho, H., Oyama, K., 2012. Endogenous Plant Cell Wall Digestion: A Key Mechanism in Insect Evolution. *Annual Review of Ecology, Evolution, and Systematics.* 43, 45-71.
- Cosgrove, D.J., 2005. Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology.* 6(11), 850-861.
- Davison, A. Blaxter, M., 2005. Ancient Origin of Glycosyl Hydrolase Family 9 Cellulase Genes. *Mol. Biol. Evol.* 22(5), 1273-1284.
- Dillon, R.J., Dillon, V.M., 2004. The Gut Bacteria of Insects: Nonpathogenic Interactions. *Annual Review of Entomology.* 49, 71-92.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research.* 32(5), 1792-1797.
- Eckert, K., Zielinski, F., Lo Leggio, L., Schneider, E., 2002. Gene cloning, sequencing, and characterization of a family 9 endoglucanase (CelA) with an unusual pattern of activity from the thermoacidophile *Alicyclobacillus acidocaldarius* ATCC27009. *Appl. Microbiol. Biotechnol.* 60, 428-436.
- Eklöf, J.M., Ruda, M.C., Brumer, H., 2012. Distinguishing xyloglucanase activity in *endo*-beta(1->4)glucanases. *Methods Enzymol.* 510, 97-120. 10.1016/B978-0-12-415931-0.00006-9
- Eyun, S.I., Wang, H., Pauchet, Y., et al., 2014. Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*). *PLoS One.* 9(4), e94052.
- Fischer, R., Ostafe, R., Twyman, R.M., 2013. Cellulases from Insects. *Adv. Biochem. Eng. Biotechnol.* 136, 51-64.
- Friedemann, K., Wipfler, B., Bradler, S., Beutel, R.G., 2012. On the head morphology of Phyllium and the phylogenetic relationships of Phasmatodea (Insecta). *Acta Zoologica.* 93(2), 184-199.

- He, X., Zhang, J., 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*. 169(2), 1157-1164.
- Hirano, N., Hasegawa, H., Nihei, S., Haruki, M., 2013. Cell-free protein synthesis and substrate specificity of full-length endoglucanase CelJ (Cel9D-Cel44A), the largest multi-enzyme subunit of the *Clostridium thermocellum* cellulosome. *FEMS Microbiol. Lett.* 344(1), 25-30.
- Hirayama, K., Watanabe, H., Tokuda, G., Kitamoto, K. Arioka, M., 2010. Purification and Characterization of Termite Endogenous β -1,4-Endoglucanases Produced in *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.* 74, 1680-1686.
- Ishurd, O., Kermagi, A., Zgheel, et al., 2004. Structural aspects of water-soluble galactomannans isolated from the seeds of *Retama raetam*. *Carbohydrate polymers*. 58(1), 41-44.
- Pastor, F.I.J., Óscar, G., Sanz-Aparicio, J., Díaz, P., 2007. Xylanases: molecular properties and applications. *Industrial enzymes*, Springer: 65-82.
- Katsuraya, K., Okuyama, K., Hatanaka, K., Oshima, R., Sato, T., Matsuzaki, K., 2003. Constitution of konjac glucomannan: chemical analysis and ^{13}C NMR spectroscopy. *Carbohydrate polymers*. 53(2), 183-189.
- Khademi, S., Guarino, L.A., Watanabe, H., Tokuda, G., Meyer, E.F., 2002. Structure of an endoglucanases from termite, *Nasutitermes takasagoensis*. *Acta Crystallographica Section D: Biological Crystallography*. 58(4), 653-659.
- Kim, S.J., Lee, C.M., Han, B.R., Kim, M.Y., Yeo, Y.S., Yoon, S.H., Koo, B.S., Jun, H.K., 2008. Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS microbiology letters*. 282(1), 44-51.
- Kirsch, R., Wielsch, N., Vogel, H., et al., 2012. Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. *BMC Genomics*. 13, 587.
- Klesov, A., 1990. Biochemistry and enzymology of cellulose hydrolysis. *Biochemistry (Mosc)*. 55(10), 1295-1318.
- Lo, N., Tokuda, G., Watanabe, H., 2011. Evolution and Function of Endogenous Termite Cellulases. *Biology of Termites: A Modern Synthesis*. Eds: D.E. Bignell, Y. Roisin, N. Lo. London, Springer: 51-67.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42(Database issue), D490-495. 10.1093/nar/gkt1178
- Misof, B., Liu, S., Meusemann, K., et al., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 346(6210), 763-767
- Morrison, M., Pope, P.B., Denman, S.E., McSweeney, C.S., 2009. Plant biomass degradation by gut microbiomes: more of the same or something new? *Curr. Opin. Biotechnol.* 20(3), 358-363.
- Ni, J., Takehara, M., Watanabe, H., 2005. Heterologous overexpression of a mutant termite cellulase gene in *Escherichia coli* by DNA shuffling of four orthologous parental cDNAs. *Biosci. Biotechnol. Biochem.* 69(9), 1711-1720.
- Ni, J., Takehara, M., Miyazawa, M., Watanabe, H., 2007. Random exchanges of non-conserved amino acid residues among four parental termite cellulases by family

- shuffling improved thermostability Protein Engineering, Design & Selection. 20(11), 535-542.
- Ni, J., Tokuda, G. 2013. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnology Advances. 31(6), 838-850.
- Oppert, C., Klingeman, W.E., Willis, J.D., Oppert, B., Jurat-Fuentes, J.L., 2010. Prospecting for cellulolytic activity in insect digestive fluids. Comparative Biochemistry and Physiology, Part B. 155, 145-154.
- Pauchet, Y., Wilkinson, P., Chauhan, R., French-Constant, R.H., 2010. Diversity of beetle genes encoding novel plant cell wall degrading enzymes. PLoS One. 5(12), e15635.
- Pauly, M., Andersen, L.N., Kauppinen, S., et al., 1999. A xyloglucan-specific endo-beta-1,4-glucanase from *Aspergillus aculeatus*: expression cloning in yeast, purification and characterization of the recombinant enzyme. Glycobiology. 9(1), 93-100.
- Potts, R., Hewitt, P., 1974. Some properties and reaction characteristics of the partially purified cellulase from the termite *Trinervitermes trinervoides* (Nasutitermitinae). Comparative Biochemistry and Physiology Part B: Comparative Biochemistry. 47(2), 327-337.
- Richardson, M.L., Reagel, P.F., Mitchell, R.F., Whitman, D.W., 2012. Opportunistic Carnivory by *Romalea* microptera (Orthoptera: Acrididae). Annals of the Entomological Society of America. 105(1), 28-35.
- Ronquist, F., Teslenko, M., van der Mark, P., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology. 61(3), 539-542.
- Scharf, M.E., Karl, Z.J., Sethi, A., Boucias, D.G., 2011. Multiple levels of synergistic collaboration in termite lignocellulose digestion. PLoS One. 6(7), e21709.
- Scrivener, A., Watanabe, H., Noda, H., 1997. Diet and carbohydrate digestion in the yellow-spotted longicorn beetle *Psacotha hilaris*. Journal of Insect Physiology. 43(11), 1039-1052.
- Shelomi, M., Lo, W.S., Kimsey, L.S., Kuo, C-H., 2013. Analysis of the gut microbiota of walking sticks (Phasmatodea). BMC Research Notes. 6(1), 368.
- Shelomi, M., Watanabe, H., Arakawa, G., 2014a. Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut. J. Insect Physiol. 60, 25-30.
- Shelomi, M., Jasper, W.C., Atallah, J., Kimsey, L.S., Johnson, B.R., 2014b. Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut. BMC Genomics. 15, 917.
- Shelomi, M., Sitepu, I.R., Boundy-Mills, K.L., Kimsey, L.S., 2015. Review of the Gross Anatomy and Microbiology of the Phasmatodea Digestive Tract. Journal of Orthoptera Research. 24(1), 29-40. doi:10.1665/034.024.0105
- Soria-Carrasco, V., Gompert, Z., Comeault, A.A., et al., 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science. 344(6185), 738-742.
- Su, X., Mackie, R.I., Cann, I.K., 2012. Biochemical and mutational analyses of a multidomain cellulase/mannanase from *Caldicellulosiruptor bescii*. Appl. Environ. Microbiol. 78(7), 2230-2240. doi: 10.1128/AEM.06814-11
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 30(12), 2725-2729.

- Terrapon, N., Li, C., Robertson, H.M., et al., 2014. Molecular traces of alternative social organization in a termite genome. *Nature communications*. 5, 3636. 10.1038/ncomms4636
- Warner, C.D., Go, R.M., García-Salinas, C., Ford, C., Reilly, P.J., 2011. Kinetic characterization of a glycoside hydrolase family 44 xyloglucanase/endoglucanase from *Ruminococcus flavefaciens* FD-1. *Enzyme Microb Technol*. 48(1), 27-32. 10.1016/j.enzmictec.2010.08.009
- Watanabe, H., Ni, J., 2013. Functional Gene Resources from Cellulose-Feeding Insects for Novel Catalysts. Biological Conversion of Biomass for Fuels and Chemicals: Exploration from Natural Utilization Systems. Eds: J. Sun, S.-Y. Ding, J. Doran-Peterson. Cambridge, UK, RSC Publishing: 156-176.
- Watanabe, H., Tokuda, G., 2001. Animal cellulases. *Cell Mol Life Sci*. 58(9), 1167-1178.
- Watanabe, H., Tokuda, G., 2010. Cellulolytic systems in insects. *Annu. Rev. Entomol*. 55, 609-632.
- Vincken, J-P., Beldman, G., Voragen, A.G., 1997. Substrate specificity of endoglucanases: what determines xyloglucanase activity? *Carbohydrate research*. 298(4), 299-310.
- Vlasenko, E., Schulein, M., Cherry, J., Xu, F., 2010. Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresour. Technol*. 101(7), 2405-2411.
- Yoshida, K., Komae, K., 2006. A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant Cell Physiol*. 47(11), 1541-1554.
- Zhang, D., Lax, A.R., Bland, J.M., Yu, J., Fedorova, N., Nierman, W.C., 2010. Hydrolysis of filter- paper cellulose to glucose by two recombinant endogenous glycosyl hydrolases of *Coptotermes formosanus*. *Insect Science*. 17(3), 245-252.
- Zhou, W., Irwin, D.C., Escovar-Kousen, J., Wilson, D.B., 2004. Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. *Biochemistry*. 43(30), 9655-9663.
- Zhou, X., Smith, J.A., Oi, F.M., Koehler, P.G., Bennett, G.W., Scharf, M.E., 2007. Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. *Gene*. 395(1), 29-39.

Figures:

Figure 1: Agarose diffusion assays of enzymatic activity in stick insect gut extracts. Enzymes were incubated overnight in wells made in gels of 0.1% substrate in 0.4% agarose (pH 5.0). Plates were stained with Congo red and enzymatic activity was detectable as clearings in the stained gel. The positive control was gut extract of the beetle *Phaedon cochleariae* (Chrysomelidae). Carboxymethylcellulose used as “cellulose.” The negative control [MilliQ water] in part confirms that microbial contamination of the gel did not occur, although sterile techniques were used in assay preparation. Blackberry leaves were used to control for background enzymes in the diet, although guts were purged of contents before use in this assay. *Aa*=*Aretaon asperimus*. *BB*=Blackberry leaves. *CMC*=Carboxymethylcellulose. *Et*=*Extatosoma tiaratum*. *Me*=*Medauroidea extradentata*. *Ps*=*Peruphasma schultei*. *Ra*=*Ramulus artemis*. *Ss*=*Sipyloidea sipylos*.

Figure 2: TLC of the *P. schultei* lumen against cellulosic and hemicellulosic substrates. The enzyme activities here thus represent those of the entire digestome of the insect, not just GH9s. Ten μ L of substrate solution were incubated for 16hr with either 10 μ L of lumen solution or 10 μ L of water (negative control), then spotted onto silica gel plates. The leftmost spot in each row is lumen alone (*PsL*), followed by pairs of substrate incubated without (-) and with (+) gut lumen. Gels were stained with 0.2% (w/v) orcinol in 9:1 methanol/sulfuric acid and developed with a heat gun. Bands indicate sugar mono- or oligomers corresponding to the glucose, xylose, mannose, and galactose #-mer markers on the right of each row, with monomers at the top and larger oligomers on the bottom. *CMC*=carboxymethylcellulose. *Filter*=Filter paper. *Glc2*=glucose dimer (cellobiose). *G#*=glucose #-mer. *Gal*=galactose. *GalM*=galactomannan. *Glc#*=glucose #-mer. *GlcM*=glucomannan. *M#*=mannose #-mer. *Man#*=mannose #-mer. *PsL*=*P. schultei* lumen. *X#*=xylose #-mer. *Xyl#*=xylose #-mer. *XyloG*=xyloglucan.

Figure 3: Evolutionary relationships of Phasmatodea GH9s to those from other Metazoans. A Bayesian-inferred phylogeny is shown comparing the predicted amino acid sequences of the Phasmatodea GH9 enzymes described here with those derived from other animals, and using bacterial GH9 sequences as an outgroup. Posterior probability support values are indicated at corresponding nodes followed by Bootstrap values (1000 replicates) derived from a maximum-likelihood analysis. The figure shows that the Phasmatodea GH9s form a monophyletic group sister to the related, polyneopteran insects: Isoptera (termites), Blattodea (cockroaches *sensu stricto*), and Orthoptera (grasshoppers, crickets, etc.). The phylogeny of the animal GH9s is similar to that of the Animalia themselves, confirming the endogenous origin of the Phasmatodea GH9s and the ancestral origin of GH9s in Metazoa (Davison & Blaxter, 2005). A larger version of the Phasmatodea cluster is used in Figure 4. Major clades are colorcoded; all insects are orange. Details of the sequences used for the analyses as well as accession numbers are provided in Table S3. Double asterisk denotes not supported.

Figure 4: Phasmatodea cellulases and their activities against several substrates. The four main branches (color coded in red, blue, green, and orange) represent gene

889 duplication events in the last common ancestor of the Phasmatodea (including *Timema*,
890 samples TCR#), while smaller groups (in shades of the above colors) represent
891 duplications within the Euphasmatodea, for seven GH9 groups in total. The outgroup is
892 *Daphnia pulex*. Values are Maximum Likelihood bootstrap values / Bayesian Analysis
893 posterior branch probability. Scale bar is the number of substitutions per site. Individual
894 enzymes were isolated, heterologously expressed, and tested for activity on different
895 substrates for only four species. X means strong activity [in terms of band strength on the
896 TLC and whether or not the substrate was completely or partially degraded, relative to
897 the activity of equal volumes of enzymes on other substrates], • means weak activity, -
898 means no activity, and a blank means the sample was not tested: either the enzyme could
899 not be heterologously expressed (TCR5) or the assay was not performed.
900 CMC=carboxymethylcellulose. FP=Filter Paper. G5=cellopentaose. G4=cellotetraose.
901 Xyl=Xylan. XG=Xyloglucan. GluM=glucomannan.

Figure 2

Figure 3

Figure 4

Phasmatodea

Euphasmatodea

AAS : *A. asperrimus*

ETI : *E. tiaratum*

MEX : *M. extradentata*

PSC : *P. schultei*

RAR : *R. artemis*

SSI : *S. sipylus*

Timematodea

TCR : *T. cristinae*

New Functions

● Xylanase

● Xyloglucanase

Enzymatic Activity

Cellulases

Hemicellulases

CMC FP G5 G4 Xyl XG GluM

0.1 substitutions per site

X = strong activity, . = weak activity, - = no activity, [blank] = not tested

Endogenous Cellulases of the Phasmatodea