Title

Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut

Journal

BMC Genomics, 15(1)

ISSN

1471-2164

Authors

Shelomi, Matan
Jasper, W Cameron
Atallah, Joel
et al.

Publication Date

2014

DOI

10.1186/1471-2164-15-917

Peer reviewed

Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut

Matan Shelomi[1,2*,†]
[*] Corresponding author
Email: mshelomi@ice.mpg.de

W Cameron Jasper[1]
Email: wcjasper@ucdavis.edu

Joel Atallah[1]
Email: joel.atallah@gmail.com

Lynn S Kimsey[1]
Email: lskimsey@ucdavis.edu

Brian R Johnson[1,†]
Email: brnjohnson@ucdavis.edu

[1] Department of Entomology and Nematology, University of California-Davis, Davis, CA 95616, USA

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

[†] Equal contributors.

Abstract

Background

Stick and leaf insects (Phasmatodea) are an exclusively leaf-feeding order of insects with no record of omnivory, unlike other "herbivorous" Polyneoptera. They represent an ideal system for investigating the adaptations necessary for obligate folivory, including plant cell wall degrading enzymes (PCWDEs). However, their physiology and internal anatomy is poorly understood, with limited genomic resources available.

Results

We *de novo* assembled transcriptomes for the anterior and posterior midguts of six diverse Phasmatodea species, with RNA-Seq on one exemplar species, *Peruphasma schultei*. The latter's assembly yielded >100,000 transcripts, with over 4000 transcripts uniquely or more highly expressed in specific midgut sections. Two to three dozen PCWDE encoding gene families, including cellulases and pectinases, were differentially expressed in the anterior midgut. These genes were also found in genomic DNA from phasmid brain tissue, suggesting endogenous production. Sequence alignments revealed catalytic sites on most PCWDE

transcripts. While most phasmid PCWDE genes showed homology with those of other insects, the pectinases were homologous to bacterial genes.

Conclusions

We identified a large and diverse PCWDE repertoire endogenous to the phasmids. If these expressed genes are translated into active enzymes, then phasmids can theoretically break plant cell walls into their monomer components independently of microbial symbionts. The differential gene expression between the two midgut sections provides the first molecular hints as to their function in living phasmids. Our work expands the resources available for industrial applications of animal-derived PCWDEs, and facilitates evolutionary analysis of lower Polyneopteran digestive enzymes, including the pectinases whose origin in Phasmatodea may have been a horizontal transfer event from bacteria.

Background

Whole transcriptome shotgun sequencing, or RNA-Seq, is a high-throughput, next-generation sequencing tool that can efficiently identify tens of thousands of functional genes in an organism or specific tissue at a given time [1,2]. This deep sequencing makes it a more attractive tool than microarrays for organisms lacking reference genomes, facilitating *de novo* transcriptome assembly [3-5]. Its high coverage is desirable when profiling transcripts in tissues of unknown function, enabling researchers to generate and/or test multiple hypotheses at once (eg: [6,7]), and in organisms potentially harboring symbiotic organisms that may or may not produce the transcripts of interest (eg: [8-10]), as RNA-Seq can simultaneously identify genes from microbes and their vectors/hosts.

Such a combination of low genome resource availability and enigmatic physiology exists in the stick and leaf insects (order Phasmatodea), or phasmids. Though common in the pre-molecular biology era through the Laboratory Stick Insect, *Carausius morosus* [11], phasmid research today is relatively limited. Few phasmids are pests of agricultural crops [11,12], though they reach plague-like abundances in temperate forests [13,14] and *C. morosus* is an invasive pest in several countries [15,16]. All life stages of all species within the order feed exclusively on leaves [17]. This obligate folivory is relatively rare: Among insects it is known only from leaf beetles (Coleoptera: Chrysomelidae), while more basal “herbivores” such as grasshoppers and crickets (Orthoptera) will quite readily scavenge vertebrate meat, engage in cannibalism, and even hunt and kill other insects [18,19]. Thus phasmids are an ideal system for studying the evolution of herbivory in the lower Polyneoptera.

Folivorous organisms benefit greatly from plant cell wall degrading enzymes (PCWDEs), a group that includes cellulases, hemicellulases, lignases, pectinases, and xylanases [17]. Once thought to be limited to microbes, endogenous (symbiont-independent) PCWDE production has since been found throughout the Animalia. In particular, cellulase (beta-1,4-endoglucanase; Enzyme Commission: 3.2.1.4) genes from the Glycoside Hydrolase family 9 (GH9) are now believed to have existed in the ancestor of all Metazoan life [20,21] as opposed to having been repeatedly acquired from microbes via horizontal gene transfer, as is thought to be origin of GH45 and GH48 cellulases in beetles [22,23]. Among insects, endogenous cellulases have been found in lower and higher termites, cockroaches, crickets, beetles [21-24], a firebrat [25], a springtail [26], and, recently, the phasmids. High cellulase activity in the anterior midguts of two phasmid species, *Eurycantha calcarata* and *Entoria*

okinawaensis, was detected, the responsible proteins isolated, and the genes encoding them sequenced. Sequence homology and antigency against an insect cellulase anti-serum supported an endogenous, Insectan origin for the enzymes [27]. Such a process is slow and predicated on the translation of PCWDE genes into enzymes active against laboratory substrates like carboxymethylcellulose or crystalline cellulose, whose specificity and selectivity are imperfect [28,29].

Whether or not phasmids contain other PCWDEs, such as the cellobiases (a type of beta-glucosidase; EC:3.2.1.21) that convert the products of cellulase into glucose monomers [24] or the pectinases (polygalacturonases; EC:3.2.1.15) that hydrolyze pectin into galacturonic acid monomers [17], was unknown. Presence of such active enzymes could explain the obligate folivory of the Phasmatodea and be a key factor in the order's evolution [30], which is itself a puzzle as the sister order to the Phasmatodea is highly debated [31,32]. Microbiological assays of the phasmid gut suggest digestion in the order is symbiont-independent [33], so any phasmid PCWDE genes are likely endogenous, yet whether the genes show homology to insect or microbe genes depends on their own evolutionary origin [23,30]. Complicating the issue is the relative lack of genetic resources for phasmids or their most closely-related orders: Orthoptera [34], Embioptera [35], and Notoptera/Xenonomia [36]. Lastly, even if phasmids have PCWDE genes, their expression is not a guarantee, nor are they necessarily expressed in the gut region where they are most active. The phasmid midgut is physically differentiated into two sections: an anterior midgut (AMG) marked by circular pleating and folding, and a posterior midgut (PMG) studded irregularly by hollow bulbs with filamentous tubules called the "appendices of the midgut" that open into the midgut lumen [37]. The appendices may have an excretory or secretory function, and the possibility exists that they produce digestive enzymes that are carried forward into the AMG via countercurrent flow [38]. In the face of all these unknowns, next-generation sequencing is the best resource for answering questions of phasmid digestive physiology efficiently and effectively.

Here we used *de novo* transcriptome assembly to identify the genes expressed in the midguts of six species of Phasmatodea from four families, while greatly increasing the publicly available genetic resources for the order. We also used RNA-Seq on one exemplar species, *Peruphasma schultei* (Pseudophasmatidae) to quantitatively compare transcript expression between the AMG and PMG, and produced a genomic DNA library from the symbiont-free phasmid brain to confirm that identified transcripts were encoded by the insect itself. Our main goal was to identify the production organ of the Phasmatodea endogenous cellulase, while simultaneously creating an inventory of expressed PCWDE and other digestive genes in phasmids and generating hypotheses on their evolutionary origins and the putative functions of the midgut sections. This study serves as a necessary preliminary for more targeted molecular work. More broadly, our transcriptomes are useful for evolutionary analyses of non-cellulase PCWDEs in insects and identifying potential genes with biotechnological applications such as in processing biofuel feedstock or improving its rheology [39,40].

Methods

Insects and microscopy

Insects used were *Peruphasma schultei* (Pseudophasmatidae), *Sipyloidea sipylus* (Diapheromeridae), *Aretaon asperrimus* (Heteropterygidae), and *Extatosoma tiaratum, Medauroidea extradentata*, and *Ramulus artemis* (Phasmatidae) cultured at room temperature in the Bohart Museum of Entomology, University of California, Davis. Phasmids were fed an *ad libitum* diet of privet (*Ligustrum* sp.) for *Peruphasma*, *Eucalyptus* for *Extatosoma*, and *Rosa* sp. for the others.

Library prep and sequencing

The RNA-Seq study of *Peruphasma schultei* made use of three biological replicates for both the anterior and the posterior midguts (AMG and PMG respectively). For each replicate, the guts of five fed, surface-sterilized, adult, male and female phasmids were removed under sterile conditions and emptied of their contents in several washes of 70% ethanol. Then the anterior and posterior sections were separately pooled and homogenized in TRIzol® Reagent. RNA was extracted according the Trizol-Plus protocol, which includes an on-column DNAase digestion step. Total RNA quality (and subsequent library quality) was checked with the Bioanalyzer 2100. Libraries were made using the Illumina TruSeq v2 kit according to the manufacturer's instructions.

Hundred base pair paired-end sequencing was performed on the HiSeq 2000 and the raw data uploaded to the NCBI SRA Database [GenBank:SRP030474]. For quality control, low quality bases and adapter contamination were removed with the fastx toolkit [41] and the cutadapt software packages [42]. FastQC [43] was used to check the final quality of reads prior to *de novo* assembly. The number of reads generated for each biological replicate is shown in Table 1.

Table 1 Total reads and trinity results for each transcriptomic or genomic library

***de novo* stick insect transcriptome assemblies**	**Reads**	**Total trinity transcripts (isotigs)**	**Total trinity components (isogroups)**	**Contig N50**
Peruphasma schultei		135622	99469	1669
Anterior midgut 1	15,578,606			
Anterior midgut 2	17,004,583			
Anterior midgut 3	20,651,269			
Total:	53,234,458			
Posterior midgut 1	17,664,733			
Posterior midgut 2	22,326,598			
Posterior midgut 3	21,932,697			
Total:	61,924,028			
Aretaon asperrimus		142181	110688	3188
Anterior midgut	57,859,873			
Posterior midgut	41,709,281			
Extatosoma tiaratum		163928	117927	1878
Anterior midgut	67,177,740			
Posterior midgut	59,190,949			
Medauroidea extradentata		130080	99465	2246
Anterior midgut	54,198,129			
Posterior midgut	49,043,590			
Ramulus artemis		169555	92260	2007
Anterior midgut	55,689,810			
Posterior midgut	59,684,645			
Sipyloidea sipylus		114125	72103	1257
Anterior midgut	47,511,044			
Genomic *P. schultei* reads (brain tissue)	46,868,237			

Total number of transcripts and components based on results of the Trinity assembler [3] with default parameters. N50 statistic is a nucleotide length**.**

For gut transcriptomes of the other five species, the same method was used as for *P. schultei* with a few changes. For each species, only one biological replicate was produced for both the anterior and the posterior midgut. This library was made of pooled midguts (all females for *E. tiaratum, M. extradentata, R. artemis*, and *S. sipylus*, and a mixture of males and females for *A. asperrimus*). RNA was successfully extracted for all tissues with the exception of the *S. sipylus* PMG, for which the extraction failed and for which no additional specimens could be obtained. RNA-extraction and quality control were as for *Peruphasma*, but libraries were made using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina kit according to the manufacturer's instructions. Sequencing was on an Illumina HiSeq 2000, and the data uploaded to the NCBI SRA Database [GenBank:SRP038202]. The numbers of reads produced for each sample are shown in Table 1.

de novo transcriptome assembly

The Trinity assembler with the default parameters was used to generate *de novo* transcriptomes for all species using quality controlled reads [3]. TopHat (v2.04) was used for aligning reads to the transcriptome [44]. HTSeq [45] was used to quantify the number of reads aligning to each transcript. Gene and isoform abundances and expression levels from the *P. schultei* RNA-Seq data were quantified using RSEM (RNA-Seq by Expectation Maximization) [46]. This program was chosen over other programs as it does not rely on reference genomes, of which there are none for the Phasmatodea. For the *P.* schultei RNA-Seq, Trinity assembly yielded 135,622 transcripts (N50 contig length =1669). Differentially expressed genes were identified using EBSeq, an R package that compares isoform expression across two or more biological conditions, in this case AMG and PMG, using a Bayesian heirarchical model [47]. Differentially expressed genes (DEGs) were those with an adjusted p-value <0.05.

Transcriptome annotation and PCWDE identification

Due to the lack of closely related species with well-annotated genomes, or even consensus as to what is the most closely related order to the Phasmatodea, we used several methods to annotate the assembled transcripts. For all *P. schultei* transcripts and the top 500 most highly expressed transcripts for the other species, we used Blast2GO's [48] tblastx program to compare each sequence to the NCBI translated nucleotide collection (nr) database, with an expect value threshold of e^{-6}. Contigs with highly significant BLAST [49] hits were mapped to the Gene Ontology (GO) database and annotated using Blast2GO with an expect value threshold of e^{-6}. InterPro annotations were performed using the Blast2GO remote connection to the InterProEBI server [48]. GO terms were modulated using ANNEX and GOSlim, using the "generic" mapping (goslim_generic.obo) available in Blast2GO (Figure 1). Potential metabolic pathways represented in the transcriptome were identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [50] database via Blast2GO (Additional files 1, 2, 3, 4, 5 and 6). Enrichment analysis (Fisher' s Exact Test via Blast2GO) was used to find enriched GO terms, with term filter value below 0.05, term filter mode "FDR," and two-tailed test options selected (Figure 2). Annotations were added to those provided by RSEM.

Figure 1 Comparisons of the GO Terms expressed in the anterior and posterior midguts of the Phasmatodea. The top 500 most expressed transcripts are seen for each of the anterior (left) and posterior (right) midguts. **A)** Aretaon asperrimus. **B)** Extatosoma tiaratum. **C)** Medauroidea extradentata. **D)** Peruphasma schultei. **E)** Ramulus artemis.

Figure 2 GO categories enriched for the most differentially expressed genes in each *Peruphasma schultei* midgut segment. Values are relative to the overall transcriptome as per Fisher's exact test. **A)** Anterior midgut (posterior midgut values in red). **B)** Posterior midgut (anterior midgut values in red).

To specifically identify PCWDE-encoding transcripts, we downloaded nucleotide sequences for representative PCWDEs from the NCBI database, selecting known, endogenous insect proteins as well as fungal, bacterial, and protozoan proteins. The query sequences from NCBI were blast-ed against the full transcriptomes after removing low-quality reads, with an expect value threshold of e^{-10}. Only transcriptome isoforms that aligned to at least 75% of the representative gene downloaded from NCBI were included in later transcript number analyses.

Amino acid alignment and phylogenetic analysis

For the putative PCWDEs, the transcript sequences from the phasmids were converted to amino acid sequences using the ExPASy online translation tool [51]. A representative sequence from each isogroup (comp#_c#) was selected based on E-value and Sim mean when compared to known enzymes in the NCBI database. The number of isogroups and isotigs (sequences) within each group is listed in Table 2, and Additional file 7: Table S7 shows these sequence names. Other known protein sequences for these enzymes were collected from the NCBI database from a diversity of organisms including bacteria, fungi, plants, other insects, nematodes, and, when available, protists, chordates, and other invertebrates (Additional file 8: Tables S8, S9, S10 and S11). The amino acid sequences were aligned using MUSCLE [52] and manually curated using Mesquite [53]. Further alignment and production of consensus sequences for clades were done using JalView [54]. These alignments were then searched for the known conserved regions for the active/catalytic sites

for each enzyme type, identified using the Catalytic Site Atlas [55], with Blast alignment to confirm their presence in the phasmid transcripts.

Table 2 Number of PCWDE isogroups and isotigs in the phasmid midgut

	# isogroups (total # isotigs)			
Species	**Cellulase**	**Pectinase**	**Cellobiase**	**Beta-1,3-glucanase**
Aretaon asperrimus	5 (16)	11 (44)	7 (14)	3 (9)
Extatosoma tiaratum	4 (14)	18 (30)	16 (27)	3 (3)
Medauroidea extradentata	7 (13)	21 (52)	12 (28)	3 (3)
Peruphasma schultei	6 (8)	7 (14)	4 (22)	3 (3)
Ramulus artemis	5 (26)	17 (70)	17 (45)	3 (6)
Sipyloidea sipylus	7 (11)	11 (36)	10 (22)	2 (4)

Data from the full midgut transcriptomes with short sequences removed. Transcripts were identified as PCWDEs based on amino acid alignment to known proteins.

MUSCLE-aligned sequences curated on Mesquite were converted to Phylip format [56]. For neighbor-joining trees, the Phylip program "seqboot" was run to make multiple datasets for bootstrapping, and the results run through "protdist" and "neighbor," then the trees combined with "consense." For parsimony trees, the "seqboot" datasets were run through "protpars" and "consense." For maximum likelihood trees, the MUSCLE-alignment file as uploaded to the CIPRES portal (www.phylo.org) [57] and run on RAxML-HPC2 on XSEDE for 1000 bootstrapping runs [58]. For Bayesian analysis, Mr.Bayes 3.2.2 [59] was run with the CIPRES datasets with 500000 generations for bootstrapping. Consensus trees were viewed and prepared for figures using FigTree 1.4.2 [60]. The Maximum Likelihood tables were chosen as the figures for this manuscript.

Testing for endogenous production of PCWDEs

The possibility existed that some transcripts came from microbial symbionts or contaminants. Table S7 shows which PCWDE encoding isogroups contained poly-adenylation signals, a feature predominantly of eukaryotic mRNA that that, unlike bacterial RNA, will pass in large amount through our cDNA synthesis method, and whose presence is used to suggest endogenicity [61-63]. We also tested the translated transcripts for the presence of eukaryote-specific signal peptides using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) [64], which is also evidence against bacterial origins for the transcripts. Such methods however cannot differentiate between enzymes produced by protozoan or fungal symbionts and those produced by insects, including insect-produced proteins whose genes were acquired from a eukaryote via horizontal gene transfer as has happened in beetles [23,30,65].

To further show that particular PCWDE genes are endogenous to the phasmid genome and not produced by gut symbionts or contaminants, we extracted DNA from non-gut tissue for next generation sequencing. Finding a gene encoding a transcript protein in the genome is a strong and frequently used indicator of endogenicity [27,66-68]. By using non-alimentary tissue we avoided aspecific amplication of microbial genes and recovered insect DNA alone, as has been done in other insects to show microbe-like genes are endogenously produced [69-71], including the first discovery of endogenously produced cellulase in an insect [72]. We dissected out the brain under sterile conditions from one *P. schultei* individual and extracted DNA using the ChargeSwitch® gDNA Mini Tissue Kit (including the RNAase digestion step). Genomic Illumina libraries for paired-end 100 bp sequencing were then produced using the Illumina Truseq kit and validated using the bioanalyzer 2100. Sequencing was conducted on the Illumina HiSeq 2000, and the data uploaded to the NCBI SRA Database

[GenBank:SRP030474]. The numbers of reads generated for the sample are shown in Table 1. We tested if the *P. schultei* cellulase and pectinase genes were endogenous in origin by mapping all genomic reads from the brain tissue back to our *de novo* midgut transcriptome assembly. We then took all the genomic reads that mapped to each PCWDE gut transcriptome gene and blasted them to the entire gut transcriptome to narrow the list of reads down to those that uniquely map to a single PCWDE gene (Table 3).

Table 3 *Peruphasma schultei* pectinase and cellulase genomic reads uniquely aligned to transcriptome isotigs

Pectinases	Genomic reads	Cellulases	Genomic reads
Comp40495_c0_seq1	16	comp22363_c1_seq1	21
Comp54109_c1_seq1	21	comp22404_c0_seq1	23
Comp55819_c1_seq1	40	comp22464_c0_seq1	12
comp55819_c1_seq2	40	comp39876_c0_seq1	20
comp55819_c1_seq3	40	comp55831_c0_seq1	34
comp56173_c8_seq1	40	comp55831_c0_seq2	33
comp56173_c8_seq2	40	comp55831_c0_seq3	40
comp56173_c8_seq3	30	comp57191_c0_seq1	39
comp56173_c8_seq4	39		
comp56691_c1_seq1	41		
comp56691_c1_seq2	20		
comp56826_c1_seq1	40		
comp56826_c2_seq1	40		
comp56826_c2_seq3	40		

These reads are therefore of endogenous genes. Counts to 40 max.

As a final test, we blasted all PCWDE transcripts to the draft genome for *Timema cristinae* (Phasmatodea: Timematidae), which was under development but available at the webpage of the Nosil Lab of Evolutionary Biology at the university of Sheffield, UK (http://nosil-lab.group.shef.ac.uk/?page_id=25). The Timematidae are considered the sister group to all other Phasmatodea [36]. While the *Timema* genome is not guaranteed to have the same genes as the species we examined, finding our transcripts within the *Timema* genome would be strong evidence that the gene is both endogenous in and ancestral to Phasmatodea.

Results

Phasmid midgut *de novo* transcriptome assemblies

From the extracted RNA libraries of the pooled AMGs or PMGs of each phasmid species we generated approx 54 million high quality, 100 bp, paired-end sequence reads (Table 1), with the exception of *Sipyloidea sipylus* (Diapheromeridae), from which we were only able to successfully extract RNA from the anterior midguts. *de novo* assembly of each midgut section's library with Trinity [3] produced ~114-170 thousand transcript contigs per species (Table 1). All reads and the final transcriptome for *Peruphasma schultei* are available under BioProject accession PRJNA221630, and for all other phasmids under PRJNA238833.

Annotation of the *P. schultei* transcriptomes

Approximately 30323 (22%) of the 135622 transcriptomic sequences had BLAST hits (Additional file 9: Table S9), most of which were homologous to sequences from other insects and arthropods (Additional file 10: Figure S1). More genes were homologous to the

red flour beetle *Tribolium castaneum* than to insects from more closely related orders, reflecting the relative dearth of available genetic information from insects in the Polyneoptera clade and the relatively recent sequencing of *Tribolium*. The high percentage of sequences with no blast hits (orthologs) is unsurprising given the lack of an annotated, sequenced Phasmatodea genome. The sequences may have represented noncoding regions, wrongly-assembled contigs, or novel genes whose significance is unknown.

Differential gene expression across the phasmid midgut

RNA-Seq analysis of *P. schultei* suggested compartmentalization of gut function (Figure 2), as found in other insects [24] and plant cell wall consuming organisms. Additional files 11 and 12 list all differentially expressed genes (DEGs) between the *P. schultei* midgut sections, defined as genes with a Posterior Probability of Differential Expression (PPDE) >0.95 [47]. We also defined genes as being highly expressed if their expression levels were 10× higher than the mean for that midgut segment. Over 4000 genes were differentially expressed in each gut section, with 2318 genes expressed only in the AMG and 1309 expressed only in the PMG. We found 318 highly expressed genes for the AMG and 648 for the PMG (Additional files 13 and 14).

Analysis of the most highly expressed sequences in each midgut section for all species further suggested compartmentalization of digestion, or at least enzyme gene expression. All species showed similar GO category profiles for each midgut section (Figure 1), with nearly 50% reduction in hydrolase gene expression in the PMG relative to the AMG. Enzyme-encoding transcripts that break down polymers at internal sites, such as serine proteases, lipases, and PCWDEs [73] were more abundant in the AMG, as were carboxylesterases transcripts and sugar hydrolases. Transcripts abundant in the PMG encoded enzymes that break down dimers and monomers, such as dipeptidases, phospholipases, and trehalase, as well as some cell membrane receptor proteins and cytochrome P450s.

Phasmatodea midgut PCWDEs

Among the sugar hydrolases were several isogroups of PCWDEs including cellulases in the GH9 family and cellobiase, which together can digest cellulose polymers completely into sugar, and the pectinase endopolygalacturonase. We also found transcripts encoding beta-1,3-glucanase (EC:3.2.1.39), a polysaccharide-degrading enzyme family known mostly from Lepidoptera larval midguts and expressed in response to feeding on a diet containing bacteria [74]. These four enzymes (cellulases, cellobiases, pectinases, and beta-1,3-glucanases) were used in the manual annotations.

Amino acid alignment and phylogenetic analysis

The phasmid cellulases aligned most closely with other Polyneopteran cellulases — including the known, active, endogenous cellulases isolated from the phasmids *Eurycantha calcarata* and *Entoria okinawaensis* [27] — as well as those of other invertebrates, tunicates, plants, and actinomycete bacteria, but not with nematode or beetle cellulases. Phasmid endoglucanases are of the GH9 family thought to be ancestral to all animal life [21], as opposed to the GH5, 45, or 48 cellulases found in nematodes and beetles [23]. The phasmid transcripts either themselves included or were homologous to transcripts including the known active sites invariant in GH9 cellulases, based on work on *Thermobifida/ Thermomonospora fusca* (PDB: 1js4) [56,57]: namely two conserved Asp's (D55, D58) functioning in catalytic

base activity and a Glu residue (E461) that functions as the catalytic acid (Figure 3). Phylogenetic analysis could not resolve domain or phylum-level relationships among the sequences tested (bootstrap values <10). Every cellulase transcript we isolated was homologous to two sequences from the *Timema* genome.

Figure 3 Sections of cellulase amino acid sequence alignments containing conserved / active site residues. Catalytic sites are identified by the grey arrows [75,76]. List of Phasmatodea sequences in Additional file 7, and all others in Additional file 7. Similar sequences from groups of related species were combined into consensus sequences. Red letters show identity with the overall consensus sequence based on conservation of physiochemical properties. Quality is the inverse likelihood of observing mutations based on the BLOSUM62 matrix [54]. *Strongylocentrotus purpuratus* and *Flavobacterium branchiophilum* are abbreviated.

The phasmid pectinase sequences aligned most closely with those of gamma proteobacteria, rather than other insects or eukaryotes. The alignment also showed Hemipteran and beetle pectinases as most similar to fungal pectinases, the latter homology already noted in the literature [30,65]. Nearly all phasmid pectinase enzymes contained the four conserved regions of the catalytic sites, based on work on *Erwinia carotovora* (PDB: 1bhe) [77]: Asn226-Thr227-Asp228, Gly248-Asp249-Asp250, Gly274-His275-Gly276, and Arg305-Ile306-Lys307 (Figure 4). An exception is the transcripts from *A. asperrimus*, whose Arg residue is replaced with a Tyrosine (Y305). Phylogenetic analysis suggested the phasmid polygalacturonases are a monophyletic group within those of the gamma proteobacteria (Figure 5).

Figure 4 Section of pectinase amino acid sequence alignments containing conserved / active site residues. See caption to Figure 3. Catalytic residues identified with grey bars [77]. *Thermoanaerobacterium thermosaccharolyticum* is abbreviated.

Figure 5 Maximum likelihood tree for the pectinases. Phylogenetic analysis of amino acid sequences made with RAxML-HPC2 on the XSEDE system [57]. Numbers are bootstrap values (1000 runs). Branch widths based on bootstrap value, branch colors based on clade. Branch lengths based on the mean number of nucleotide substitutions per site (Scale Bar =0.9). All Phasmatodea sequences (Additional file 7) were a monophyletic group among the gamma proteobacteria. *Thermoanaerobacterium thermosaccharolyticum* is abbreviated.

The possibility exists that the pectinase transcripts came from gut bacteria or contaminants rather than phasmid genes. However, many of the pectinase transcripts had poly-A tails, as did those of other PCWDEs (†in Table S7), and all the PCWDE transcripts that were not truncated at the 5' end had eukaryotic-specific signal peptides. This includes transcripts that contained complete open reading frames (* in Table S7) as well as those that were 3' truncated. Each pectinase-encoding contig also had multiple (in most cases very many) matching genomic reads from brain tissue that uniquely aligned to them (Table 3). The same matching genomic reads could be found for cellulases, which have been demonstrated to be endogenously produced in phasmids [27] and are endogenously produced in many other insects [17] and metazoans [21]. However, none of the pectinase transcripts had homologues in the *Timema* genome.

The phasmid beta-glucosidases/cellobiases were in the GH1 family and aligned most with those of other insects. The phasmid transcripts mostly had the conserved residues of beta-

glucosidases, including the catalytic sites, based on work with white clover, *Trifolium repens* (PDB: 1cbg) [78]: Arg75, His119, Asn163, Glu164, Asn306, Tyr308, Glu378, and Trp420 (Figure 6). Phylogenetic analysis suggested the phasmid cellobiases are nearly all monophyletic, except a strongly-supported clade consisting of one isogroup from each species but *S. sipylus* (Figure 7). Analysis could not determine interclass relationships among insect beta-glucosidases, but suggested the enzyme existed in the common ancestor of the Insecta. Every beta-glucosidase transcript we isolated was homologous to four sequences from the *Timema* genome.

Figure 6 Sections of cellobiase amino acid sequence alignments containing conserved / active site residues. See caption to Figure 3. Catalytic and conserved residues identified with grey arrows [78].

Figure 7 Maximum likelihood tree for the beta-glucosidases. Phylogenetic analysis of amino acid sequences made with RAxML-HPC2 on the XSEDE system [57]. Numbers are bootstrap values (1000 runs). Branch widths based on bootstrap value, branch colors based on clade. Branch lengths based on the mean number of nucleotide substitutions per site (Scale Bar =0.7). The Phasmatodea sequences (Additional file 7) formed two strongly supported monophyletic group with weak relationships to other insect groups.

The phasmid beta-1,3-glucanases were in the GH16 family and aligned most closely with other insect enzymes (Figure 8), however this is a relatively recently described enzyme with few recorded sequences in the literature or NCBI database. This is the first known record of endogenous beta-1,3-glucanase in the Polyneoptera. The phasmid beta-1,3-glucanases could be divided into four clear, monophyletic groups (Figure 9) with no more than one representative isogroup per each of the six phasmid species. Each group differed in their homology with the known consensus pattern for catalytically active beta-1,3-glucanases based on work on *Bacillus licheniformis* [79]: E-[LIV]-D-[LIVF]-x(0,1)-E-x(2)-[GQ]-[KRNF]-x-[PSTA] (Figure 8: 342–353). One group of six sequences had 11/12 amino acids conserved, a group of four had 10/12, another group of six had 8/12, and the final group consisting of one *A. asperrimus* sequence had 6/12. The last amino acid in the *Bacillus* region was not conserved among any phasmids, nor is it conserved among the Lepidoptera sequences, which were also 11/12, or many other organism sequences sampled. Every beta-1,3-glucanase transcript we isolated had six to eight homologues in the *Timema* genome.

Figure 8 Sections of beta-1,3-glucanase amino acid sequence alignments containing conserved / active site residues. See caption to Figure 3. Conserved, twelve amino acid region identified with grey bar [79]. The four groups of Phasmatodea gene are separated by black lines.

Figure 9 Maximum likelihood tree for the beta-1,3-glucanases. Phylogenetic analysis of amino acid sequences made with RAxML-HPC2 on the XSEDE system [57]. Numbers are bootstrap values (1000 runs). Branch widths based on bootstrap value, branch colors based on clade. Branch lengths based on the mean number of nucleotide substitutions per site (Scale Bar =0.7). The Phasmatodea sequences formed four strongly supported groups.

Discussion

Using high coverage sequencing of RNA expressed in their midguts, we were able to produce high quality transcriptomes of several Phasmatodea species. This new data doubles the genera of phasmids with publicly available genetic resources on the NCBI databases, increasing the amount of annotated genes available for future work not only on Phasmatodea, but also on the Polyneoptera in general. Covering six species in four families, while drawing from the draft genome of a seventh species in a fifth family, the data suggests the differential expression and enzyme gene diversity of the phasmid midgut sections is mostly conserved throughout the order. Our findings will serve as a reference set for studying phasmid digestion and a jumping point for future proteomic and biochemical assays. The abundance of PCWDE isogroups in phasmids is relatively high, and the diversity of PCWDE types is comparable to those in certain leaf beetles (Chrysomelidae) like *Phaedon cochleariae* [69,71] or wood-boring beetles (Cerambycidae) like *Anoplophora glabripennis* [80]. The current record is likely *Diabrotica virgifera virgifera* (Chrysomelidae) with seventy-eight genes putatively encoding proteins from the same four enzyme classes studied here [55]. As Phasmatodea and Chrysomelidae are among the few insect groups to be exclusively folivorous, a possible correlation exists between that dietary niche and a diverse PCWDE complement.

For *de novo* transcriptomes, assemblers such as Trinity often cannot differentiate between homologous genes and isoforms or allelic variants of the same gene. They can potentially overestimate the number of isotigs (single or groups of contigs that should each constitute one splice variant) within an isogroup (all isotigs for one gene, identified by Trinity's output as comp#_c#) [81]. Combined with relatively low genetic resource availability for closely related insects and relatively high representation of species like the aforementioned beetles, we cannot be certain whether phasmids express more or fewer PCWDEs than the average herbivorous insect. In addition to using programs like RSEM designed to reduce such errors [46], comparing the number of reads mapping to a locus on the genome can be used to infer the true isoform number and account for inflation [82]. That most *P. schultei* transcriptome sequences (contigs) had more matching genome reads (Table 3) than their corresponding isogroup has members (Table 2) suggests that our contig numbers represent true isoforms within each isogroup, rather than an overestimation due to mis-assembly [82,83]. These phasmid isoforms may reflect multiple gene copies or alternatively spliced genes, either case suggesting a diverse complement of proteins working together to fully digest multiple varieties of carbohydrate polymer. a highly derived genetic capacity for plant cell wall breakdown [84]. However, because some isotigs were truncated at the 3' or 5' end, the possibility exists that certain transcripts represent different ends of a single gene. Future work using RACE-PCR from primers based on the transcripts identified here would produce full-length cDNA sequences that will determine which transcripts represent unique genes and which are fragments, Such genes could then be expressed into insect cell cultures for use in downstream enzymatic activity assays [85].

Previous research has confirmed that the endogenous cellulase genes we demonstrated are most highly expressed in the anterior midgut are also most highly active in the anterior midgut [27], making it the site of both cellulase translation and action. The physical structure of the AMG supports this hypothesis: the pleating and folding serves to greatly increase the available surface area of the AMG while slowing down the transit speed of food, increasing the amount of time and space available for cellulase enzymes to hydrolyze ingested plant material. Cellulase activity falls to nearly nothing in the PMG, tracking with cellulase gene

expression. If we extend the results of cellulases to those of the other PCWDEs, then we hypothesize that phasmid digestive enzymes are active in the same region of the gut where they are expressed, making the pleated AMG the site of primary plant cell wall and polymer digestion and the PMG the site of secondary digestion of smaller oligomers at most.

We also hypothesize that phasmids can fully digest cellulose into glucose, as they have the two enzymes necessary to do so, and also actively degrade pectin into galacturonic acid. Such digestive abilities could explain how phasmids survive on otherwise uncommon, obligately folivorous diets: by fully breaking down plant cell walls into assimilatable nutrients rather than just degrading the walls to access the nutrient-rich cytoplasm within. As transcriptomics only demonstrates gene expression, not translation or activity, these hypotheses cannot be confirmed with this data alone. However, the fact that phasmids have active cellulases [27] and the presence of the relevant catalytic residues on the phasmid cellulase, pectinase, and cellobiase transcripts (Figures 3,4,6) and some beta-1,3-glucanase transcripts (Figure 8) suggests the transcripts code for functional enzymes, supporting the hypothesis that these enzymes are indeed actively degrading plant cell walls. We have thus provided the necessary preliminary work justifying biochemical and proteomic assays into cellobiase, pectinase, and beta-glucanase activity in the phasmid gut.

Phasmid pectinases are all endopolygalacturonases in the GH28 group, known in insects only from the beetles and the Hemiptera, but they show homology and align to those from gamma proteobacteria (Figure 4). Pectinase genes were also absent in the *Timema* genome. Our pectinase transcripts may have come from a bacterial symbiont. The successful mapping of all *P. schultei* pectinase and cellulase transcripts to genes in the *P. schultei* genome, the presence of eukaryote-specific poly-A tails and signal peptides on phasmid transcripts with complete open reading frames, previous studies with *P. schultei* and *R. artemis* suggesting their digestion is symbiont independent [33], and the absence of characteristic paunches for microbial fermentation [37], all tentatively suggest these pectinases are encoded in the phasmids own genome and not produced by gut microbes. However, the possibility remains that the samples were contaminated by a non-symbiotic microbe: either a rare bacteria that poly-adenylates its RNA or a fungal symbiont that acquired a bacterial gene via horizontal transfer.

A more parsimonious hypothesis is that the phasmid pectinase gene was acquired through horizontal transfer from a bacterial ancestor, much as the beetle pectinases are thought to have been acquired though horizontal transfer from an Ascomycete fungus [23,30,65], or as leaf beetle xylanases may have also transferred from a gamma proteobacteria [69]. The absence of the genes in *Timema* would suggest either that the transfer event occurred after the split between the Timematidae and the other Phasmatodea, or that the pectinase genes are ancestral to both and lost in *Timema*. Lastly, the similarity between phasmid and bacterial sequences could simply be an artefact of the over-representation of microbial and dearth of animal pectinases in the NCBI database at this time [86]. Using long range or RACE PCR to clone entire genes from phasmid genomic DNA and get introns would conclusively demonstrate whether or not the pectinases are endogenously produced or not, and such work on all six species studied here is underway.

Whether pectinase genes exist in other Polyneoptera remains to be seen, but would help determine when the horizontal transfer event could have occurred. Failure to find endogenous pectinases in closely related insects would mean the transfer occurred in an early Phasmatodea ancestor: a development that would have expanded the digestive abilities of the

order and may have played a significant role in their evolution of obligate folivory. PCWDE diversity could also be correlated to the development of the longer and larger body sizes of the Euphasmatodea. Broader, multi-phyla, phylogenetic analysis for pectinolytic enzyme genes in the Animalia can answer this question [35].

Phasmid cellobiases are GH1, as are those of other insects who produce them endogenously like the higher termites [87]. Such species can break down cellobiose independently, unlike the lower termites that have symbiotic microorganisms to produce their cellobiases for them. The beta-1,3-glucanases are GH16, similar to those found in Lepidoptera [74], however the prevalence of this recently described gene in animals has not been sufficiently examined. A greater sampling of this gene's presence in other animal, fungal, and bacterial species, as well as biochemical studies to determine the conserved catalytic residues for the protein, are needed before ordinal-level hypotheses can be made for the enzyme's evolutionary history. So far, the animal enzymes appear most closely related, and we hypothesize that at least three distinct beta-1,3-glucanase gene families existed in an early Phasmatodea ancestor, including the ancestor of the Timematidae.

Our work promotes phasmids as a potentially high-value source of novel PCWDEs for biotechnological applications [70]. Cellulases and pectinases are highly sought after by the biofuel industry to degrade feedstock into the monomers later converted into fuel, or to improve the flow rate of the material by reducing the amount of solid matter [39]. Pectinases are also used in the production of coffee, tea, and juice, and in waste-water treatment [88]. Phasmid PCWDEs could be introduced into bacteria or fungi like *Trichoderma reesei*, for industrial-scale enzyme production or direct use in bioreactors for wastewater treatment or biofuel production [89].

Our *de novo* midgut transcriptomes enabled us to survey all expressed PCWDEs of the Phasmatodea at once and identify conserved catalytic domains, justifying downstream translation and activity level analyses. A benefit of this is system is the increased speed and efficiency compared to the converse [90]: running chemical assays to identify enzyme activity, using proteomics to identify the amino acid sequence of isolated enzymes, and working backwards from there to design a primer for the enzyme-encoding gene and hope it exists within the target organism’s genome itself and not a symbiont or contaminant [91]. Another benefit is that transcriptomics can reveal genes useful for phylogenetic analysis but that are not translated or whose proteins are modified post-translation such that the standard biochemical tests do not detect their function. An associated drawback is that expressed genes are not necessarily translated into active proteins, nor are they necessarily active at the site of expression [92]. However, when a reference genome is not available, a transcriptome can provide large sets of potential genes for study and, combined with genomic data, can determine whether or not they are endogenous to the target organism, which cannot be determined by homology alone. *de novo* transcriptome assembly combined with RNA-Seq is a powerful tool for suggesting putative functions for unknown tissues in understudied organisms and directions for future study.

Conclusions

The folivorous Phasmatodea are an ideal system to study the evolution of obligate herbivory, yet a paucity of genetic resources and poorly understood basic biology impede such work. Using RNA-Seq, we demonstrated a diversity of plant cell wall degrading enzymes expressed

differentially in the anterior section of the phasmid midgut. Of these, the cellulases, cellobiases, and beta-1,3-glucanases are likely all encoded in the insect's own genome, as are the pectinases, though we could not definitively rule out a microbial source for the latter. Such an abundance of endogenous enzymes was not expected from the Polyneoptera, raising important questions on their evolutionary history. The efficiency by which our *de novo* transcriptomes generated new genomic resources and hypotheses for future research on Polyneopteran digestion demonstrate the power of such methods to analyze organisms lacking sequenced genomes. Our findings strongly encourage expanding the searches for PCWDEs, most notably the pectinases and beta-1,3-glucanases, into other, lower Polyneopteran insects.

Availability of supporting data

All reads and sequence files described in the manuscript are available under BioProject accessions PRJNA221630 for *P. schultei* and PRJNA238833 for the other phasmids.

Abbreviations

AMG, Anterior midgut; CIPRES, Cyberinfrastructure for phylogenetic research; DEG, Differentially expressed genes; EC, Enzyme commission; GH, Glycoside hydrolase; GO, Gene ontology; HEG, Highly expressed gene; KEGG, Kyoto encyclopedia of genes and genomes; NCBI, National center for biotechnology information; PCWDE, Plant cell wall degrading enzyme; PDB, Protein data bank; PMG, Posterior midgut; PPDE, Posterior probability of differential expression; RPKM, Reads per kilobase per million; RSEM, RNA-Seq by expectation maximization; SRA, Sequence read archive; XSEDE, Extreme science and engineering discovery environment

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors MS and BRJ conceived the study and contributed equally to the work. LSK and BRJ provided resources and expertise. MS and LSK reared the insects. WCJ, JA, and BRJ carried out the RNA Seq and *de novo* transcriptome assembly. MS, WCJ, JA, and BRJ analyzed and interpreted data. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Dr. Steve Heydon and the Bohart Museum of Entomology staff and volunteers for insect rearing, to the Dr. Patrik Nosil Lab of the University of Sheffield, UK, for their work on the *Timema* genome and making it available online, and to Dr. Yannick Pauchet of the Max Planck Institute of Chemical Ecoogy in Jena, Germany, for general advising. MS was supported by the National Science Foundation (USA) Graduate Research Fellowship under Grant No. 1148897, the UC Davis & Humanities Graduate Research Fellowship in Entomology for 2012–13 and 2013–14, and the McBeth Memorial Scholarship. The research

was supported by funds from the University of California at Davis. Thanks also to the editors and reviewers who contributed for their feedback.

References

1. Ekblom R, Galindo J: **Applications of next generation sequencing in molecular ecology of non-model organisms.** *Heredity (Edinb)* 2011, **107**(1)**:**1–15.

2. Metzker ML: **Sequencing technologies - the next generation.** *Nat Rev Genet* 2010, **11**(1)**:**31–46.

3. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29**(7)**:**644–652.

4. Xia Z, Xu H, Zhai J, Li D, Luo H, He C, Huang X: **RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*.** *Plant Mol Biol* 2011, **77**(3)**:**299–308.

5. Francis WR, Christianson LM, Kiko R, Powers ML, Shaner NC, Haddock SH D: **A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly.** *BMC Genomics* 2013, **14:**167.

6. Poelchau MF, Reynolds JA, Denlinger DL, Elsik CG, Armbruster PA: **A de novo transcriptome of the Asian tiger mosquito, Aedes albopictus, to identify candidate transcripts for diapause preparation.** *BMC Genomics* 2011, **12:**619.

7. Hull JJ, Geib SM, Fabrick JA, Brent CS: **Sequencing and de novo assembly of the western tarnished plant bug (Lygus hesperus) transcriptome.** *PLoS One* 2013, **8**(1)**:**e55105.

8. McCarthy CB, Santini MS, Pimenta PF, Diambra LA: **First comparative transcriptomic analysis of wild adult male and female *Lutzomyia longipalpis*, vector of visceral leishmaniasis.** *PLoS One* 2013, **8**(3)**:**e58645.

9. Yazawa T, Kawahigashi H, Matsumoto T, Mizuno H: **Simultaneous transcriptome analysis of Sorghum and Bipolaris sorghicola by using RNA-seq in combination with de novo transcriptome assembly.** *PLoS One* 2013, **8**(4)**:**e62460.

10. Lehnert EM, Mouchka ME, Burriesci MS, Gallo ND, Schwarz JA, Pringle JR: **Extensive differences in gene expression between symbiotic and aposymbiotic Cnidarians.** *G3 (Bethesda)* 2014, **4**(2)**:**277–295.

11. Bedford GO: **Biology and ecology of the phasmatodea.** *Annu Rev Entomol* 1978, **23:**125–149.

12. Kasenene JM: **Forest association and phenology of wild coffee in Kibale National Park, Uganda.** *Afr J Ecol* 1998, **36:**241–250.

13. Readshaw JL: **Phasmatid outbreaks revisiting?** *Aust J Zool* 1990, **38:**343–346.

14. Jurskis V, Turner J: **Eucalypt dieback in eastern Australia: a simple model.** *Aust For* 2002, **65**(2)**:**87–98.

15. Headrick D, Wilen CA: **Indian Walking Stick**. In *Pest Notes*, Volume 74157. ; 2011:1–3.

16. Borges PA, Reut M, Ponte NB, Quartau JA, Fletcher M, Sousa AB, Pollet M, Soares AO, Marcelino J, Rego C: **New records of exotic spiders and insects to the Azores, and new data on recently introduced species.** *Arquipélago Life and Marine Sciences* 2013, **30:**57–70.

17. Calderón-Cortés N, Quesada M, Watanabe H, Cano-Camacho H, Oyama K: **Endogenous plant cell wall digestion: a key mechanism in insect evolution.** *Annu Rev Ecol Evol Syst* 2012, **43:**45–71.

18. Whitman DW, Blum MS, Slansky F Jr: **Carnivory in Phytophagous Insects**. In *Functional Dynamics of Phytophagous Insects.* Edited by Ananthakrishnan TN. New Delhi: Oxford & IBH Publishing Co. Pvt. Ltd; 1994:161–205.

19. Whitman DW, Richardson ML: **Necrophagy in grasshoppers: *Taeniopoda eques* feeds on mammal varrion.** *J Orthopt Res* 2010, **19**(2)**:**377–380.

20. Lo N, Watanabe H, Sugimura M: **Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals.** *Proc Biol Sci* 2003, **270**(Suppl 1)**:**S69–72.

21. Davison A, Blaxter M: **Ancient origin of glycosyl hydrolase family 9 cellulase genes.** *Mol Biol Evol* 2005, **22**(5)**:**1273–1284.

22. Calderón-Cortés N, Watanabe H, Cano-Camacho H, Zavala-Páramo G, Quesada M: **cDNA cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle *Oncideres albomarginata chamela* (Cerambycidae).** *Insect Mol Biol* 2010, **19**(3)**:**323–336.

23. Eyun SI, Wang H, Pauchet Y, Ffrench-Constant RH, Benson AK, Valencia-Jimenez A, Moriyama EN, Siegfried BD: **Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*).** *PLoS One* 2014, **9**(4)**:**e94052.

24. Fischer R, Ostafe R, Twyman RM: **Cellulases from insects.** *Adv Biochem Eng Biotechnol* 2013, **136:**51–64.

25. Treves DS, Martin MM: **Cellulose digestion in primitive hexapods: effect of ingested antibiotics on gut microbial populations and gut cellulase levels in the firebrat, *Thermobia domestica* (Zygentoma, Lepismatidae).** *J Chem Ecol* 1994, **20**(8)**:**2003–2020.

26. Hong SM, Sung HS, Kang MH, Kim C-G, Lee Y-H, Kim D-J, Lee JM, Kusakabe T: **Characterization of cryptopygus antarcticus endo-β-1, 4-glucanase from bombyx Mori expression systems.** *Mol Biotechnol* 2014, **56**(**10**)**:**878–889.

27. Shelomi M, Watanabe H, Arakawa G: **Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut.** *J Insect Physiol* 2014, **60:**25–30.

28. Watanabe H, Tokuda G: **Cellulolytic systems in insects.** *Annu Rev Entomol* 2010, **55:**609–632.

29. Willis JD, Oppert C, Jurat-Fuentes JL: **Methods for discovery and characterization of cellulolytic enzymes from insects.** *Insect Sci* 2010, **17:**184–198.

30. Kirsch R, Gramzow L, Theissen G, Siegfried BD, Ffrench-Constant RH, Heckel DG, Pauchet Y: **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* 2014, **52C:**33–50.

31. Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK: **Advances in insect phylogeny at the dawn of the postgenomic era.** *Annu Rev Entomol* 2012, **57:**449–468.

32. Letsch H, Simon S: **Insect phylogenomics: new insights on the relationships of lower neopteran orders (Polyneoptera).** *Syst Entomol* 2013, **38**(4)**:**783–793.

33. Shelomi M, Lo WS, Kimsey LS, Kuo CH: **Analysis of the gut microbiota of walking sticks (Phasmatodea).** *BMC Res Notes* 2013, **6**(1)**:**368.

34. Terry MD, Whiting MF: **Mantophasmatodea and phylogeny of the lower neopterous insects.** *Cladistics* 2005, **21:**240–257.

35. Letsch HO, Meusemann K, Wipfler B, Schutte K, Beutel R, Misof B: **Insect phylogenomics: results, problems and the impact of matrix composition.** *Proc Biol Sci* 2012, **279**(1741)**:**3282–3290.

36. Plazzi F, Ricci A, Passamonti M: **The mitochondrial genome of *Bacillus* stick insects (Phasmatodea) and the phylogeny of orthopteroid insects.** *Mol Phylogenet Evol* 2011, **58**(2)**:**304–316.

37. Shelomi M, Kimsey LS: **Vital staining of the stick insect digestive system identifies appendices of the midgut as novel system of excretion.** *J Morphol* 2014, **275**(6)**:**623–633.

38. Monteiro EC, Tamaki FK, Terra WR, Ribeiro AF: **The digestive system of the "stick bug" Cladomorphus phyllinus (Phasmida, Phasmatidae): a morphological, physiological and biochemical analysis.** *Arthropod Struct Dev* 2014, **43**(2)**:**123–134.

39. Geddes CC, Nieves IU, Ingram LO: **Advances in ethanol production.** *Curr Opin Biotechnol* 2011, **22**(3)**:**312–319.

40. Jurat-Fuentes JL, Oppert C, Klingeman W, Oppert B: **Identification and Characterization of Insect Cellulolytic Systems for Plant Biomass Degradation**. In *Sun Grant Initiative: Southeastern Regional Center*. Knoxville, TN: University of Tennessee; 2011:24. http://sungrant.tennessee.edu/NR/rdonlyres/BD78756A-539B-4001-88B8-6C801F81E9A2/2987/FuentesFINAL.pdf.

41. Hannon G: **FASTX-toolkit.** 2012, [http://hannonlabcshledu/fastx_toolkit].

42. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet journal* 2011, **17**(1)**:**10–12.

43. Andrews S: *FastQC: A Quality Control Tool for High Throughput Sequence Data.* Cambridge, UK: Babraham Bioinformatics; 2010.

44. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**(9)**:**1105–1111.

45. Anders S: **HTSeq: analysing high-throughput sequencing data with python.** 2010, [http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html].

46. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC bioinformatics* 2011, **12:**323.

47. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C: **EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments.** *Bioinformatics* 2013, **29**(8)**:**1035–1043.

48. Conesa A, Gotz S: **Blast2GO: a comprehensive suite for functional analysis in plant genomics.** *Int J Plant Genom* 2008, **2008:**619832.

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3)**:**403–410.

50. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y: **KEGG for linking genomes to life and the environment.** *Nucleic Acids Res* 2008, **36**(Database issue)**:**D480–D484.

51. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H: **ExPASy: SIB bioinformatics resource portal.** *Nucleic Acids Res* 2012, **40**(W1)**:**W597–W603.

52. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**(5)**:**1792–1797.

53. Maddison WP, Maddison DR: **Mesquite: a modular system for evolutionary analysis. Version 2.75.** 2011, [http://mesquiteproject.org].

54. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: **Jalview Version 2–a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**(9)**:**1189–1191.

55. Furnham N, Holliday GL, de Beer TA, Jacobsen JO, Pearson WR, Thornton JM: **The catalytic site atlas 2.0: cataloging catalytic sites and residues identified in enzymes.** *Nucleic Acids Res* 2014, **42**(Database issue)**:**D485–489.

56. Felsenstein J: *PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author.* Seattle: Department of Genome Sciences, University of Washington; 2005.

57. Miller MA, Pfeiffer W, Schwartz T: **Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees**. In *Gateway Computing Environments Workshop (GCE).* ; 2010:1–8. IEEE.

58. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**(21)**:**2688–2690.

59. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.** *Syst Biol* 2012, **61**(3)**:**539–542.

60. Rambaut A: **FigTree, a graphical viewer of phylogenetic trees.** 2007, [http://treebioedacuk/software/figtree].

61. Perl A, Rosenblatt JD, Chen IS, DiVincenzo JP, Bever R, Poiesz J, Abraham GN: **Detection and cloning of new HTLV-related endogenous sequences in man.** *Nucleic Acids Res* 1989, **17**(17)**:**6841–6854.

62. Edmonds M: **A history of poly A sequences: from formation to factors to function.** *Prog Nucleic Acid Res Mol Biol* 2002, **71:**285–389.

63. Davis R, Shi Y: **The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation.** *J Zhejiang Univ Sci B* 2014, **15**(5)**:**429–437.

64. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8**(10)**:**785–786.

65. Shen Z, Denton M, Mutti N, Pappan K, Kanost MR, Reese JC, Reeck GR: **Polygalacturonase from *Sitophilus oryzae*: possible horizontal transfer of a pectinase gene from fungi to weevils.** *J Insect Sci* 2003, **3:**24.

66. Kim N, Choo YM, Lee KS, Hong SJ, Seol KY, Je YH, Sohn HD, Jin BR: **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* 2008, **150:**368–376.

67. Pauchet Y, Saski CA, Feltus FA, Luyten I, Quesneville H, Heckel DG: **Studying the organization of genes encoding plant cell wall degrading enzymes in Chrysomela tremula provides insights into a leaf beetle genome.** *Insect Mol Biol* 2014, **23**(3)**:**286–300.

68. Chauhan R, Jones R, Wilkinson P, Pauchet Y: **Cytochrome P450-encoding genes from the Heliconius genome as candidates for cyanogenesis.** *Insect Mol Biol* 2013, **22**(5)**:**532–540.

69. Pauchet Y, Heckel DG: **The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer.** *Proc Biol Sci* 2013, **280**(1763)**:**20131021.

70. Busconi M, Berzolla A, Chiappini E: **Preliminary data on cellulase encoding genes in the xylophagous beetle, *Hylotrupes bajulus* (Linnaeus).** *Int Biodeterior Biodegradation* 2014, **86:**92–95.

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

72. Watanabe H, Tokuda G: **Animal cellulases.** *Cell Mol Life Sci* 2001, **58**(9)**:**1167–1178.

73. Ortego F: **Physiological Adaptations of the Insect Gut to Herbivory**. In *Arthropod-Plant Interactions: Novel Insights and Approaches for IPM*, Volume 14. Edited by Smagghe G, Diaz I. Dordrecht, NY: Springer; 2012:75–88.

74. Pauchet Y, Freitak D, Heidel-Fischer HM, Heckel DG, Vogel H: **Immunity or digestion: glucanase activity in a glucan-binding protein family from Lepidoptera.** *J Biol Chem* 2009, **284**(4)**:**2214–2224.

75. Sakon J, Irwin D, Wilson DB, Karplus PA: **Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*.** *Nat Struct Biol* 1997, **4**(10)**:**810–818.

76. Zhou W, Irwin DC, Escovar-Kousen J, Wilson DB: **Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes.** *Biochemistry* 2004, **43**(30)**:**9655–9663.

77. Pickersgill R, Smith D, Worboys K, Jenkins J: **Crystal structure of polygalacturonase from Erwinia carotovora ssp. carotovora.** *J Biol Chem* 1998, **273**(38)**:**24660–24664.

78. Barrett T, Suresh CG, Tolley SP, Dodson EJ, Hughes MA: **The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase.** *Structure* 1995, **3**(9)**:**951–960.

79. Juncosa M, Pons J, Dot T, Querol E, Planas A: **Identification of active site carboxylic residues in Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase by site-directed mutagenesis.** *J Biol Chem* 1994, **269**(20)**:**14530–14535.

80. Scully ED, Hoover K, Carlson JE, Tien M, Geib SM: **Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle.** *BMC Genomics* 2013, **14:**850.

81. O'Neil ST, Emrich SJ: **Assessing *De Novo* transcriptome assembly metrics for consistency and utility.** *BMC Genomics* 2013, **14:**465.

82. Jiang H, Wong WH: **Statistical inferences for isoform expression in RNA-Seq.** *Bioinformatics* 2009, **25**(8)**:**1026–1032.

83. Vijay N, Poelstra JW, Kunstner A, Wolf JB: **Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments.** *Mol Ecol* 2013, **22**(3)**:**620–634.

84. Scharf ME, Karl ZJ, Sethi A, Sen R, Raychoudhury R, Boucias DG: **Defining host-symbiont collaboration in termite lignocellulose digestion.** *Commun Integr Biol* 2011, **4**(6)**:**761–763.

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

86. Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH: **Diversity of beetle genes encoding novel plant cell wall degrading enzymes.** *PLoS One* 2010, **5**(12)**:**e15635.

87. Tokuda G, Watanabe H, Hojo M, Fujita A, Makiya H, Miyagi M, Arakawa G, Arioka M: **Cellulolytic environment in the midgut of the wood-feeding higher termite *Nasutitermes takasagoensis*.** *J Insect Physiol* 2012, **58**(1)**:**147–154.

88. Kashyap DR, Vohra PK, Chopra S, Tewari R: **Applications of pectinases in the commercial sector: a review.** *Bioresour Technol* 2001, **77**(3)**:**215–227.

89. Dashtban M, Schraft H, Qin W: **Fungal bioconversion of lignocellulosic residues; opportunities & perspectives.** *Int J Biol Sci* 2009, **5**(6)**:**578.

90. Li LL, McCorkle SR, Monchy S, Taghavi S, van der Lelie D: **Bioprospecting metagenomes: glycosyl hydrolases for converting biomass.** *Biotechnol Biofuels* 2009, **2:**10.

91. Oppert C, Klingeman WE, Willis JD, Oppert B, Jurat-Fuentes JL: **Prospecting for cellulolytic activity in insect digestive fluids.** *Comp Biochem Physiol Biochem Mol Biol* 2010, **155:**145–154.

92. Góngora-Castillo E, Buell CR: **Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence.** *Nat Prod Rep* 2013, **30**(4)**:**490–500.

Additional files

Additional_file_1 as XLS

Additional file 1: Table S1. KEGG Table for the most highly expressed genes of the *Aretaon asperrimus* midgut. Top 500 most highly expressed genes in each of the anterior midgut (AMG) and posterior midgut (PMG) used.

Additional_file_2 as XLS

Additional file 2: Table S2. KEGG Table for the most highly expressed genes of the *Extatosoma tiaratum* midgut. Top 500 most highly expressed genes in each of the anterior midgut (AMG) and posterior midgut (PMG) used.

Additional_file_3 as XLS

Additional file 3: Table S3. KEGG Table for the most highly expressed genes of the

Medauroidea extradentata midgut. Top 500 most highly expressed genes in each of the anterior midgut (AMG) and posterior midgut (PMG) used.

Additional_file_4 as XLS

Additional file 4: Table S4. KEGG Table for the most highly expressed genes of the *Peruphasma schultei* midgut. Top 500 most highly expressed genes in each of the anterior midgut (AMG) and posterior midgut (PMG) used.

Additional_file_5 as XLS

Additional file 5: Table S5. KEGG Table for the most highly expressed genes of the *Ramulus artemis* midgut. Top 500 most highly expressed genes in each of the anterior midgut (AMG) and posterior midgut (PMG) used.

Additional_file_6 as XLS

Additional file 6: Table S6. KEGG Table for the most highly expressed genes of the *Sipyloidea sipylus* midgut. Top 500 most highly expressed genes in each of the anterior midgut (AMG) only.

Additional_file_7 as XLS

Additional file 7: Table S7. Representative isotigs (sequences) for each PCWDE isogroup for the phasmatodea. Data is from the full midgut transcriptomes with short sequences removed. Transcripts were identified as PCWDEs based on amino acid alignment to known proteins from the NCBI database. * = The sequence contained a complete open-reading frame. Other sequences could have been truncated at the 5' or 3' or both. † = The sequence or another in that isogroup (_c#) had a poly-A tail.

Additional_file_8 as XLS

Additional file 8 Table S8 – Species and NCBI Accession No's for cellulase (Beta-1,4-endoglucanase) proteins compared to phasmid proteins. Sequences with * were not included in the alignment figure (Figure 4) due to poor or absent overlap with other sequences at that region. **Table S9** – Species and NCBI Accession No's for pectinase (polygalacturonase) proteins compared to phasmid proteins. Sequences with * were not included in the alignment figure (Figure 5) due to poor or absent overlap with other sequences at that region. **Table S10** – Species and NCBI Accession No's for cellobiase (beta-glucosidase) proteins compared to phasmid proteins. Sequences with * were not included in the alignment figure (Figure 7) due to poor or absent overlap with other sequences at that region. **Table S11** – Species and NCBI Accession No's for beta-1,3-glucanase proteins compared to phasmid proteins. Sequences with * were not included in the alignment figure (Figure 9) due to poor or absent overlap with other sequences at that region.

Additional_file_9 as TXT

Additional file 9: Table S9. Number of *Peruphasma schultei* midgut transcriptome sequences with successful Blast search, mapping, and annotation.

Additional_file_10 as PDF

Additional file 10: Figure S1. Species distribution for top-hit Blast results of *P. schultei* midgut transcriptome.

Additional_file_11 as XLS

Additional file 11: Table S12. The most differentially expressed genes (DEGs) in the *P. schultei* AMG. Includes genes found in both or only one tissue type. Means measured in RPKM. PPDE = Posterior Probability of Differential Expression. Annotations made with Blast2GO.

Additional_file_12 as XLS

Additional file 12: Table S13. The most differentially expressed genes (DEGs) in the *P. schultei* PMG. Includes genes found in both or only one tissue type. Means measured in RPKM. PPDE = Posterior Probability of Differential Expression. Annotations made with Blast2GO.

Additional_file_13 as XLS

Additional file 13: Table S14. The most highly expressed genes of the *P. schultei* AMG. Identified as genes with expression levels ten times greater than the mean for that section. Means measured in RPKM. Annotations made with Blast2GO.

Additional_file_14 as XLS

Additional file 14: Table S15. The most highly expressed genes of the *P. schultei* PMG. Identified as genes with expression levels ten times greater than the mean for that section. Means measured in RPKM. Annotations made with Blast2GO.

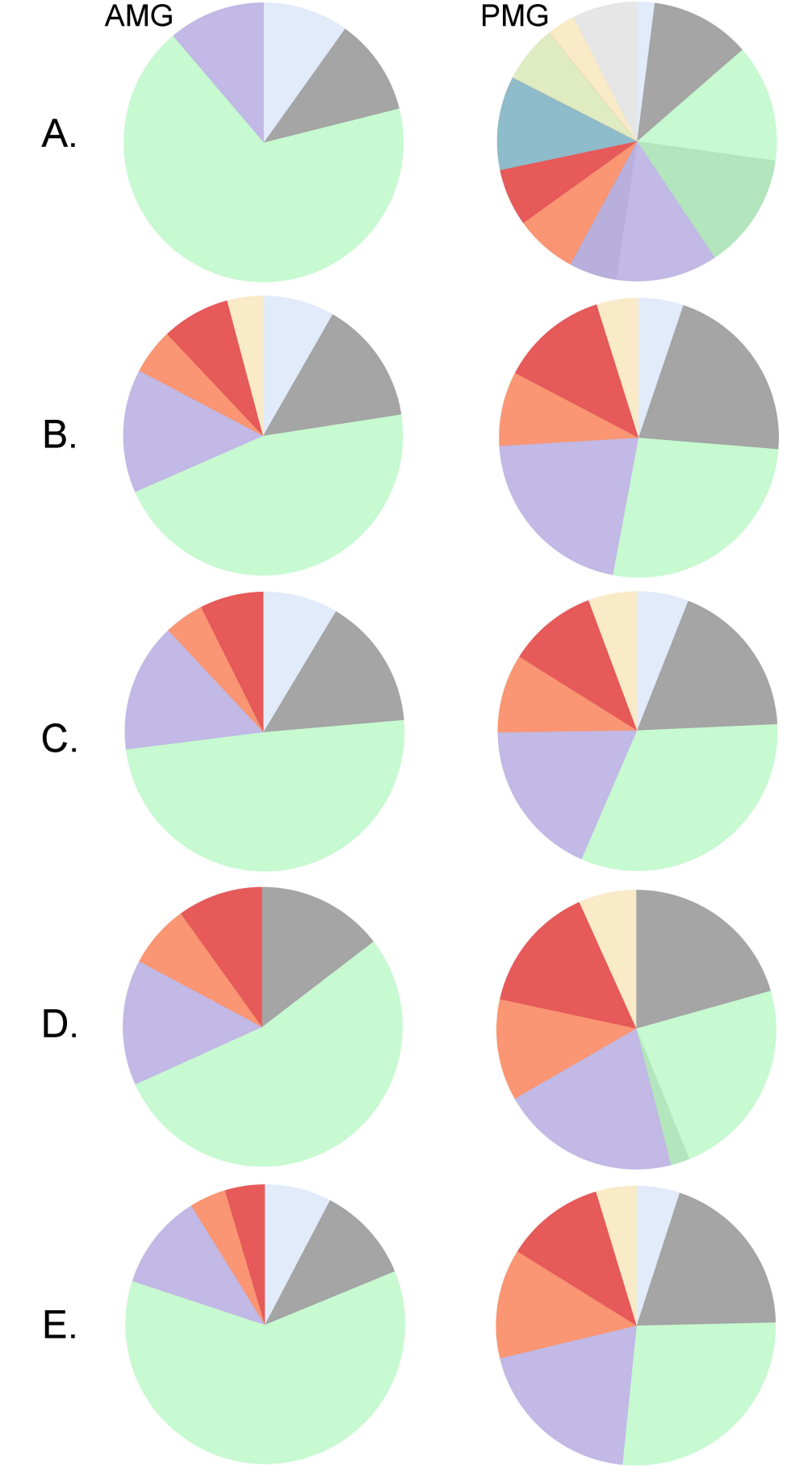
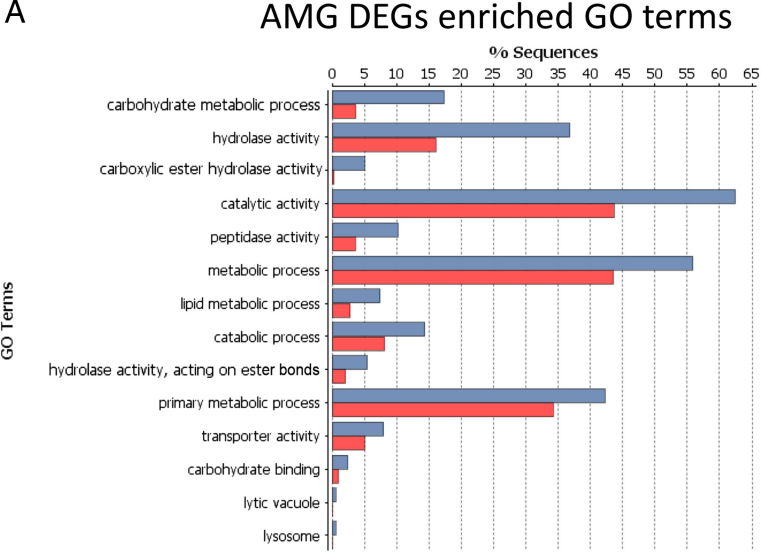

AMG
PMG
A.
B.
C.
D.
E.
carbohydrate binding
hydrolase activity
organic cyclic compound binding
protein binding
structural constitutent of ribosome
transferase activity
heterocyclic compound binding
ion binding
oxidoreductase activity
small molecule binding
substrate-specific transporter
transmembrane transporter

Figure 1

A


AMG DEGs enriched GO terms
% Sequences
0 5 10 15 20 25 30 35 40 45 50 55 60 65
GO Terms
carbohydrate metabolic process
hydrolase activity
carboxylic ester hydrolase activity
catalytic activity
peptidase activity
metabolic process
lipid metabolic process
catabolic process
hydrolase activity, acting on ester bonds
primary metabolic process
transporter activity
carbohydrate binding
lytic vacuole
lysosome

B

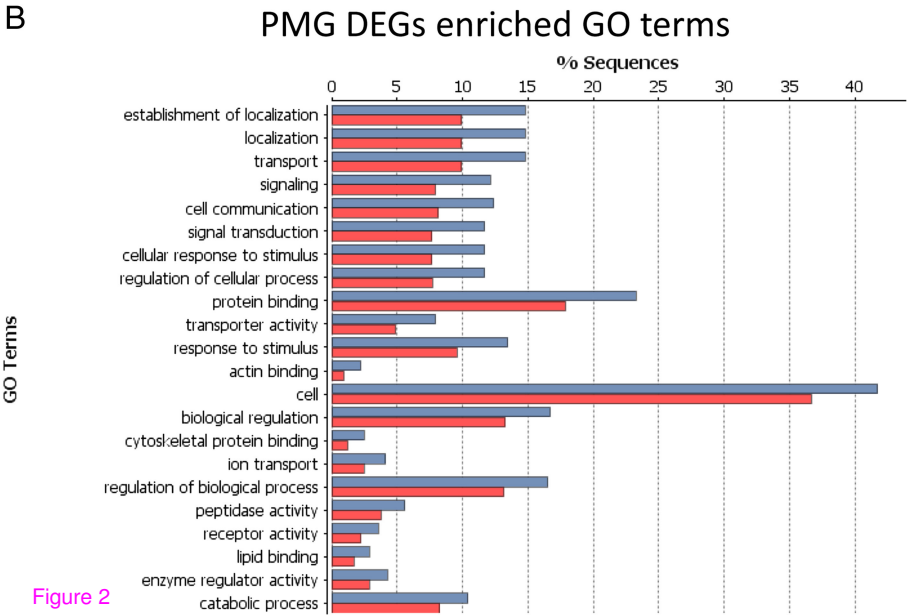

PMG DEGs enriched GO terms
% Sequences
0 5 10 15 20 25 30 35 40
GO Terms
establishment of localization
localization
transport
signaling
cell communication
signal transduction
cellular response to stimulus
regulation of cellular process
protein binding
transporter activity
response to stimulus
actin binding
cell
biological regulation
cytoskeletal protein binding
ion transport
regulation of biological process
peptidase activity
receptor activity
lipid binding
enzyme regulator activity
catabolic process

Figure 2

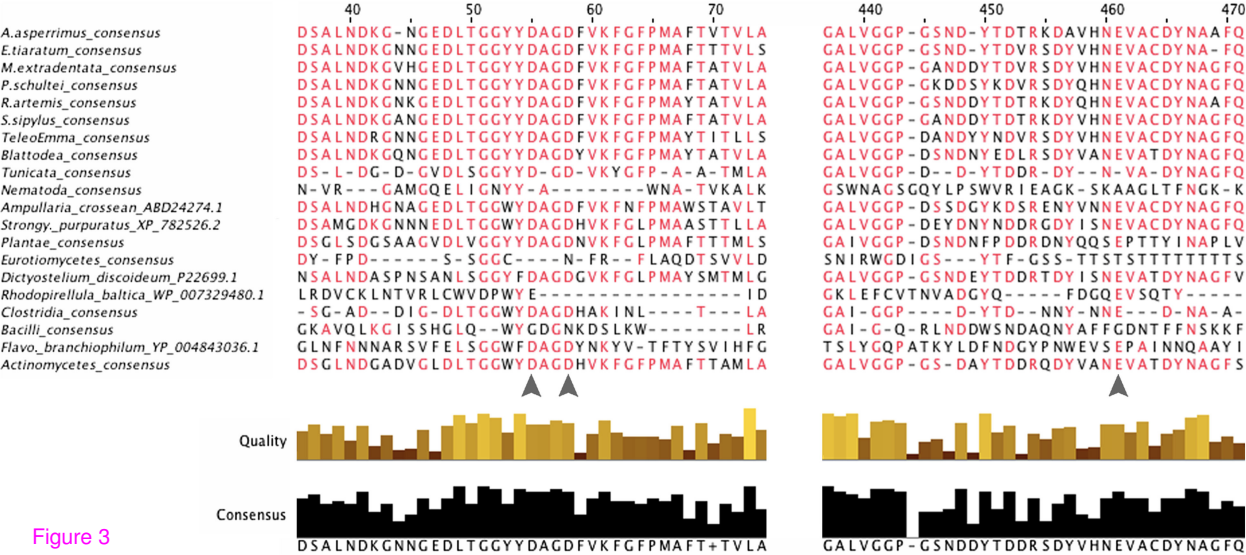

40
50
60
70
440
450
460
470
A.asperrimus_consensus
E.tiaratum_consensus
M.extradentata_consensus
P.schultei_consensus
R.artemis_consensus
S.sipylus_consensus
TeleoEmma_consensus
Blattodea_consensus
Tunicata_consensus
Nematoda_consensus
Ampullaria_crossean_ABD24274.1
Strongy._purpuratus_XP_782526.2
Plantae_consensus
Eurotiomycetes_consensus
Dictyostelium_discoideum_P22699.1
Rhodopirellula_baltica_WP_007329480.1
Clostridia_consensus
Bacilli_consensus
Flavo._branchiphilum_YP_004843036.1
Actinomycetes_consensus
Quality
Consensus
DSALNDKGNNGEDLTGGYYDAGDFVKFGFPMAFT+TVLA
GALVGGP-GSNDDYTDDRSDYVHNEVACDYNAGFQ

Figure 3

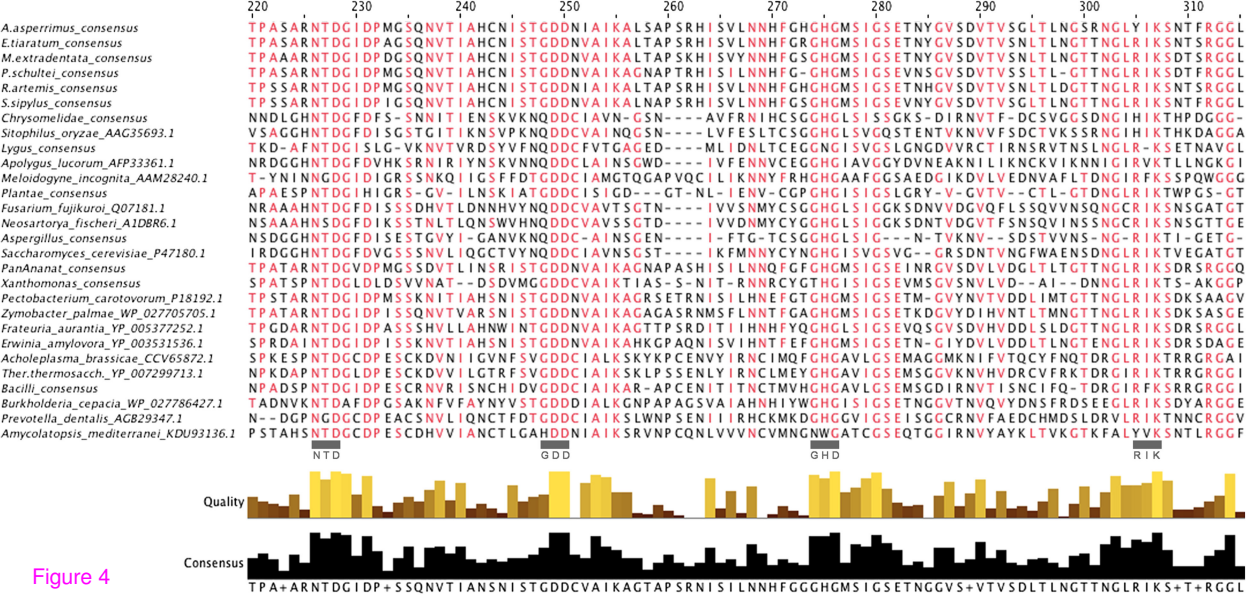

220 230 240 250 260 270 280 290 300 310
A.asperrimus_consensus
E.tiaratum_consensus
M.extradentata_consensus
P.schultei_consensus
R.artemis_consensus
S.sipylus_consensus
Chrysomelidae_consensus
Sitophilus_oryzae_AAG35693.1
Lygus_consensus
Apolygus_lucorum_AFP33361.1
Meloidogyne_incognita_AAM28240.1
Plantae_consensus
Fusarium_fujikuroi_Q07181.1
Neosartorya_fischeri_A1DBR6.1
Aspergillus_consensus
Saccharomyces_cerevisiae_P47180.1
PanAnanat_consensus
Xanthomonas_consensus
Pectobacterium_carotovorum_P18192.1
Zymobacter_palmae_WP_027705705.1
Frateuria_aurantia_YP_005377252.1
Erwinia_amylovora_YP_003531536.1
Acholeplasma_brassicae_CCV65872.1
Ther.thermosacch._YP_007299713.1
Bacilli_consensus
Burkholderia_cepacia_WP_027786427.1
Prevotella_dentalis_ACB29347.1
Amycolatopsis_mediterranei_KDU93136.1
NTD
GDD
GHD
RIK
Quality
Consensus

Figure 4

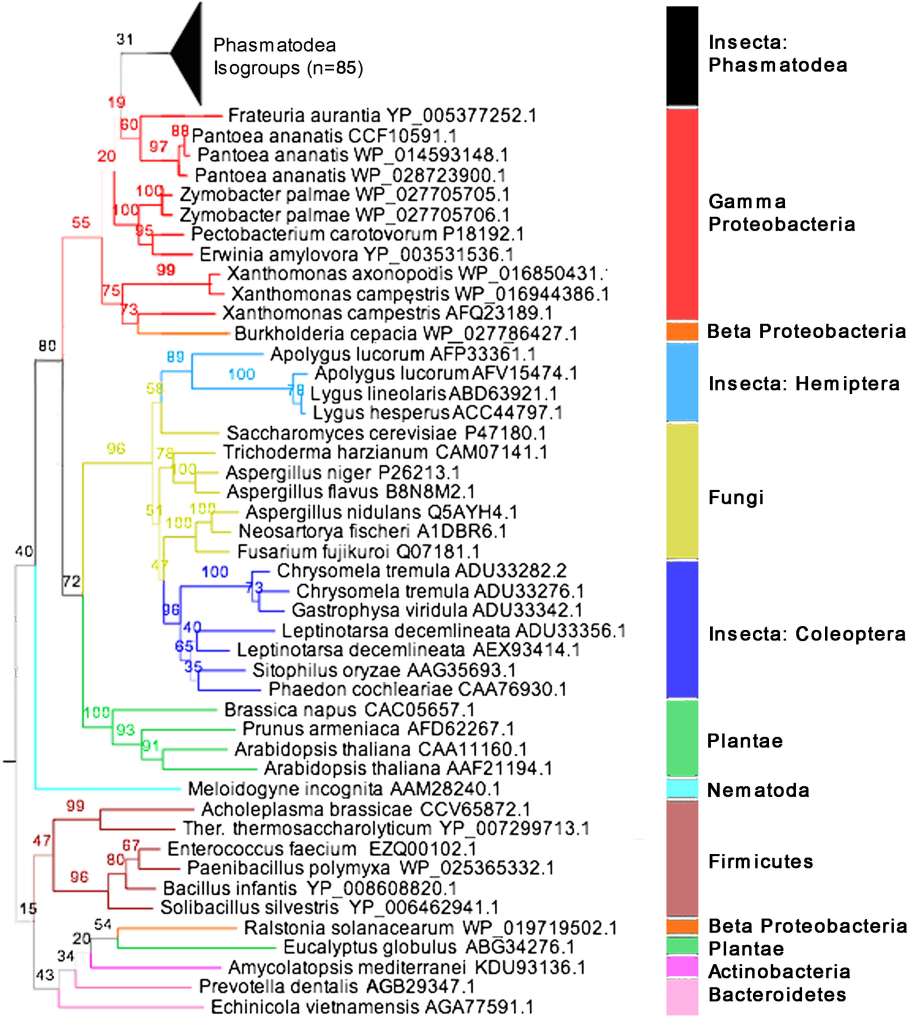

Phasmatodea
Isogroups (n=85)
Frateuria aurantia YP_005377252.1
Pantoea ananatis CCF10591.1
Pantoea ananatis WP_014593148.1
Pantoea ananatis WP_028723900.1
Zymobacter palmae WP_027705705.1
Zymobacter palmae WP_027705706.1
Pectobacterium carotovorum P18192.1
Erwinia amylovora YP_003531536.1
Xanthomonas axonopodis WP_016850431.
Xanthomonas campestris WP_016944386.1
Xanthomonas campestris AFQ23189.1
Burkholderia cepacia WP_027786427.1
Apolygus lucorum AFP33361.1
Apolygus lucorum AFV15474.1
Lygus lineolaris ABD63921.1
Lygus hesperus ACC44797.1
Saccharomyces cerevisiae P47180.1
Trichoderma harzianum CAM07141.1
Aspergillus niger P26213.1
Aspergillus flavus B8N8M2.1
Aspergillus nidulans Q5AYH4.1
Neosartorya fischeri A1DBR6.1
Fusarium fujikuroi Q07181.1
Chrysomela tremula ADU33282.2
Chrysomela tremula ADU33276.1
Gastrophysa viridula ADU33342.1
Leptinotarsa decemlineata ADU33356.1
Leptinotarsa decemlineata AEX93414.1
Sitophilus oryzae AAG35693.1
Phaedon cochleariae CAA76930.1
Brassica napus CAC05657.1
Prunus armeniaca AFD62267.1
Arabidopsis thaliana CAA11160.1
Arabidopsis thaliana AAF21194.1
Meloidogyne incognita AAM28240.1
Acholeplasma brassicae CCV65872.1
Ther. thermosaccharolyticum YP_007299713.1
Enterococcus faecium EZQ00102.1
Paenibacillus polymyxa WP_025365332.1
Bacillus infantis YP_008608820.1
Solibacillus silvestris YP_006462941.1
Ralstonia solanacearum WP_019719502.1
Eucalyptus globulus ABG34276.1
Amycolatopsis mediterranei KDU93136.1
Prevotella dentalis AGB29347.1
Echinicola vietnamensis AGA77591.1
Insecta: Phasmatodea
Gamma Proteobacteria
Beta Proteobacteria
Insecta: Hemiptera
Fungi
Insecta: Coleoptera
Plantae
Nematoda
Firmicutes
Beta Proteobacteria
Plantae
Actinobacteria
Bacteroidetes
0.9

Figure 5

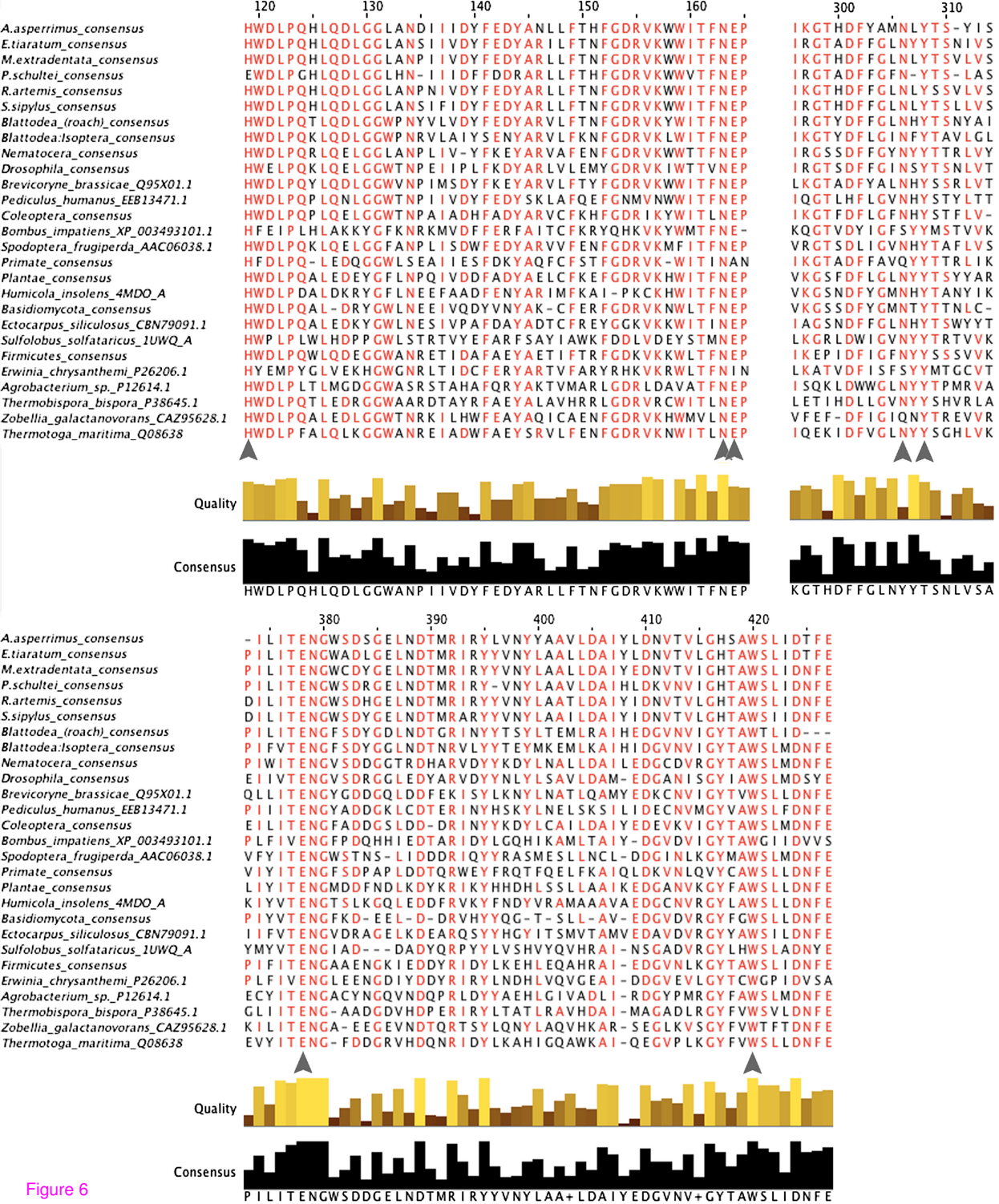

120
130
140
150
160
300
310
A.asperrimus_consensus
E.tiaratum_consensus
M.extradentata_consensus
P.schultei_consensus
R.artemis_consensus
S.sipylus_consensus
Blattodea_(roach)_consensus
Blattodea:Isoptera_consensus
Nematocera_consensus
Drosophila_consensus
Brevicoryne_brassicae_Q95X01.1
Pediculus_humanus_EEB13471.1
Coleoptera_consensus
Bombus_impatiens_XP_003493101.1
Spodoptera_frugiperda_AAC06038.1
Primate_consensus
Plantae_consensus
Humicola_insolens_4MDO_A
Basidiomycota_consensus
Ectocarpus_siliculosus_CBN79091.1
Sulfolobus_solfataricus_1UWQ_A
Firmicutes_consensus
Erwinia_chrysanthemi_P26206.1
Agrobacterium_sp._P12614.1
Thermobispora_bispora_P38645.1
Zobellia_galactanovorans_CAZ95628.1
Thermotoga_maritima_Q08638
Quality
Consensus
HWDLPQHLQDLGGWANPIIVDYFEDYARLLFTNFGDRVKWWITFNEP
KGTHDFFGLNYYTSNLVSA
380
390
400
410
420
A.asperrimus_consensus
E.tiaratum_consensus
M.extradentata_consensus
P.schultei_consensus
R.artemis_consensus
S.sipylus_consensus
Blattodea_(roach)_consensus
Blattodea:Isoptera_consensus
Nematocera_consensus
Drosophila_consensus
Brevicoryne_brassicae_Q95X01.1
Pediculus_humanus_EEB13471.1
Coleoptera_consensus
Bombus_impatiens_XP_003493101.1
Spodoptera_frugiperda_AAC06038.1
Primate_consensus
Plantae_consensus
Humicola_insolens_4MDO_A
Basidiomycota_consensus
Ectocarpus_siliculosus_CBN79091.1
Sulfolobus_solfataricus_1UWQ_A
Firmicutes_consensus
Erwinia_chrysanthemi_P26206.1
Agrobacterium_sp._P12614.1
Thermobispora_bispora_P38645.1
Zobellia_galactanovorans_CAZ95628.1
Thermotoga_maritima_Q08638
Quality
Consensus
PILITENGWSDDGELNDTMRIRYYVNYLAA+LDAIYEDGVNV+GYTAWSLIDNFE

Figure 6

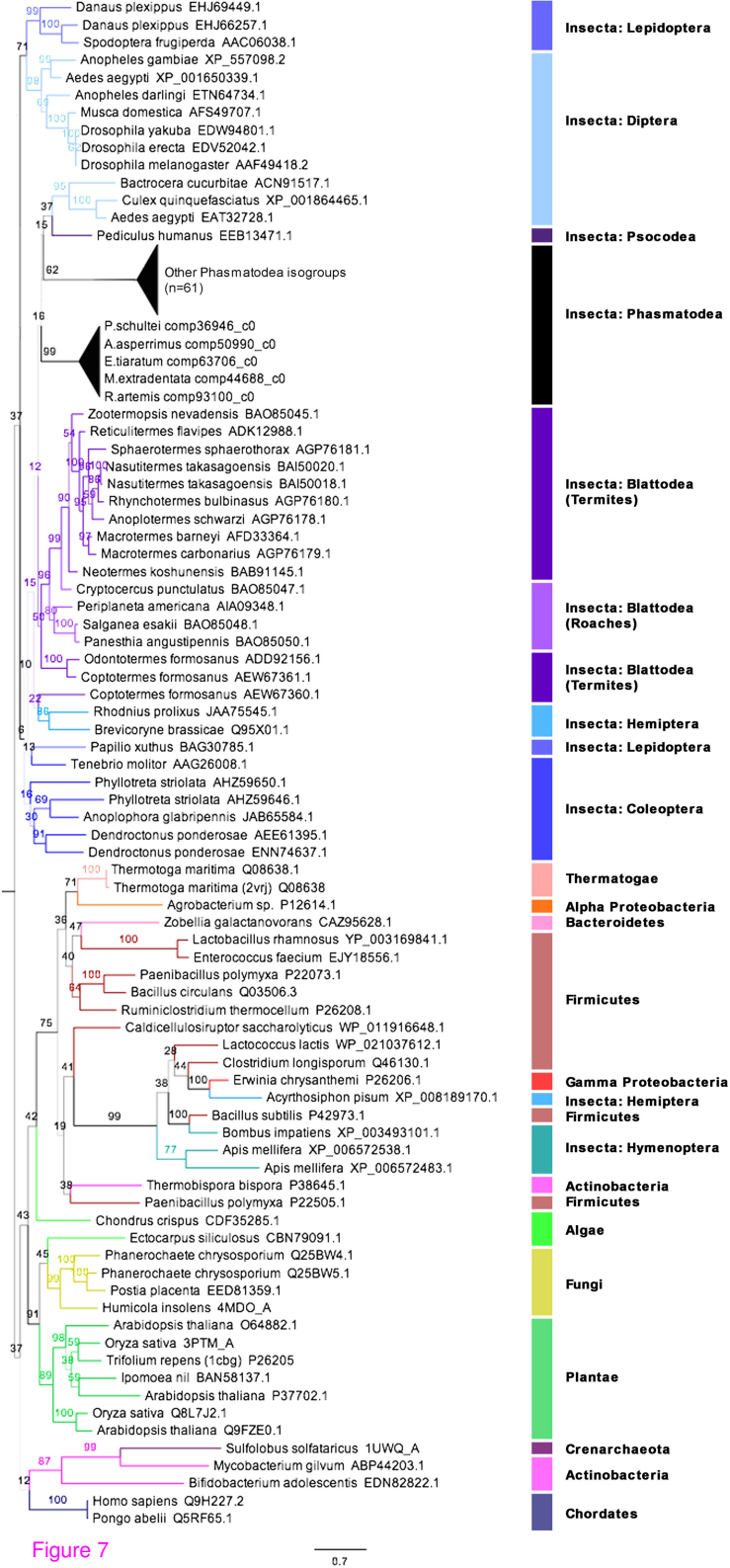

Danaus plexippus EHJ69449.1
Danaus plexippus EHJ66257.1
Spodoptera frugiperda AAC06038.1
Anopheles gambiae XP_557098.2
Aedes aegypti XP_001650339.1
Anopheles darlingi ETN64734.1
Musca domestica AFS49707.1
Drosophila yakuba EDW94801.1
Drosophila erecta EDV52042.1
Drosophila melanogaster AAF49418.2
Bactrocera cucurbitae ACN91517.1
Culex quinquefasciatus XP_001864465.1
Aedes aegypti EAT32728.1
Pediculus humanus EEB13471.1
Other Phasmatodea isogroups (n=61)
P.schultei comp36946_c0
A.asperrimus comp50990_c0
E.tiaratum comp63706_c0
M.extradentata comp44688_c0
R.artemis comp93100_c0
Zootermopsis nevadensis BAO85045.1
Reticulitermes flavipes ADK12988.1
Sphaerotermes sphaerothorax AGP76181.1
Nasutitermes takasagoensis BAI50020.1
Nasutitermes takasagoensis BAI50018.1
Rhynchotermes bulbinasus AGP76180.1
Anoplotermes schwarzi AGP76178.1
Macrotermes barneyi AFD33364.1
Macrotermes carbonarius AGP76179.1
Neotermes koshunensis BAB91145.1
Cryptocercus punctulatus BAO85047.1
Periplaneta americana AIA09348.1
Salganea esakii BAO85048.1
Panesthia angustipennis BAO85050.1
Odontotermes formosanus ADD92156.1
Coptotermes formosanus AEW67361.1
Coptotermes formosanus AEW67360.1
Rhodnius prolixus JAA75545.1
Brevicoryne brassicae Q95X01.1
Papilio xuthus BAG30785.1
Tenebrio molitor AAG26008.1
Phyllotreta striolata AHZ59650.1
Phyllotreta striolata AHZ59646.1
Anoplophora glabripennis JAB65584.1
Dendroctonus ponderosae AEE61395.1
Dendroctonus ponderosae ENN74637.1
Thermotoga maritima Q08638.1
Thermotoga maritima (2vrj) Q08638
Agrobacterium sp. P12614.1
Zobellia galactanovorans CAZ95628.1
Lactobacillus rhamnosus YP_003169841.1
Enterococcus faecium EJY18556.1
Paenibacillus polymyxa P22073.1
Bacillus circulans Q03506.3
Ruminiclostridium thermocellum P26208.1
Caldicellulosiruptor saccharolyticus WP_011916648.1
Lactococcus lactis WP_021037612.1
Clostridium longisporum Q46130.1
Erwinia chrysanthemi P26206.1
Acyrthosiphon pisum XP_008189170.1
Bacillus subtilis P42973.1
Bombus impatiens XP_003493101.1
Apis mellifera XP_006572538.1
Apis mellifera XP_006572483.1
Thermobispora bispora P38645.1
Paenibacillus polymyxa P22505.1
Chondrus crispus CDF35285.1
Ectocarpus siliculosus CBN79091.1
Phanerochaete chrysosporium Q25BW4.1
Phanerochaete chrysosporium Q25BW5.1
Postia placenta EED81359.1
Humicola insolens 4MDO_A
Arabidopsis thaliana O64882.1
Oryza sativa 3PTM_A
Trifolium repens (1cbg) P26205
Ipomoea nil BAN58137.1
Arabidopsis thaliana P37702.1
Oryza sativa Q8L7J2.1
Arabidopsis thaliana Q9FZE0.1
Sulfolobus solfataricus 1UWQ_A
Mycobacterium gilvum ABP44203.1
Bifidobacterium adolescentis EDN82822.1
Homo sapiens Q9H227.2
Pongo abelii Q5RF65.1
Insecta: Lepidoptera
Insecta: Diptera
Insecta: Psocodea
Insecta: Phasmatodea
Insecta: Blattodea (Termites)
Insecta: Blattodea (Roaches)
Insecta: Blattodea (Termites)
Insecta: Hemiptera
Insecta: Lepidoptera
Insecta: Coleoptera
Thermatogae
Alpha Proteobacteria
Bacteroidetes
Firmicutes
Gamma Proteobacteria
Insecta: Hemiptera
Firmicutes
Insecta: Hymenoptera
Actinobacteria
Firmicutes
Algae
Fungi
Plantae
Crenarchaeota
Actinobacteria
Chordates
0.7

Figure 7

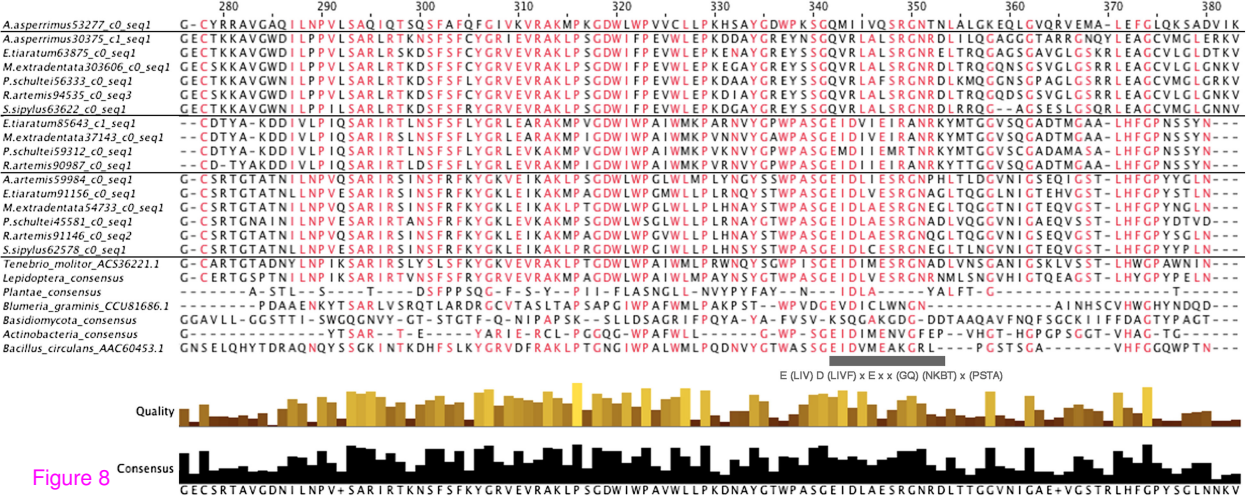

280 290 300 310 320 330 340 350 360 370 380
A.asperrimus53277_c0_seq1
A.asperrimus30375_c1_seq1
E.tiaratum63875_c0_seq1
M.extradentata303606_c0_seq1
P.schultei56333_c0_seq1
R.artemis94535_c0_seq3
S.sipylus63622_c0_seq1
E.tiaratum85643_c1_seq1
M.extradentata37143_c0_seq1
P.schultei59312_c0_seq1
R.artemis90987_c0_seq1
A.artemis59984_c0_seq1
E.tiaratum91156_c0_seq1
M.extradentata54733_c0_seq1
P.schultei45581_c0_seq1
R.artemis91146_c0_seq2
S.sipylus62578_c0_seq1
Tenebrio_molitor_ACS36221.1
Lepidoptera_consensus
Plantae_consensus
Blumeria_graminis_CCU81686.1
Basidiomycota_consensus
Actinobacteria_consensus
Bacillus_circulans_AAC60453.1
E (LIV) D (LIVF) x E x x (GQ) (NKBT) x (PSTA)
Quality
Consensus
GECSRTAVGDNILNPV+SARIRTKNSFSFKYGRVEVRAKLPSGDWIWPAVWLLPKDNAYGTWPASGEIDLAESRGNRDLTTGGVNIGAE+VGSTRLHFGPYSGLNNKV

Figure 8

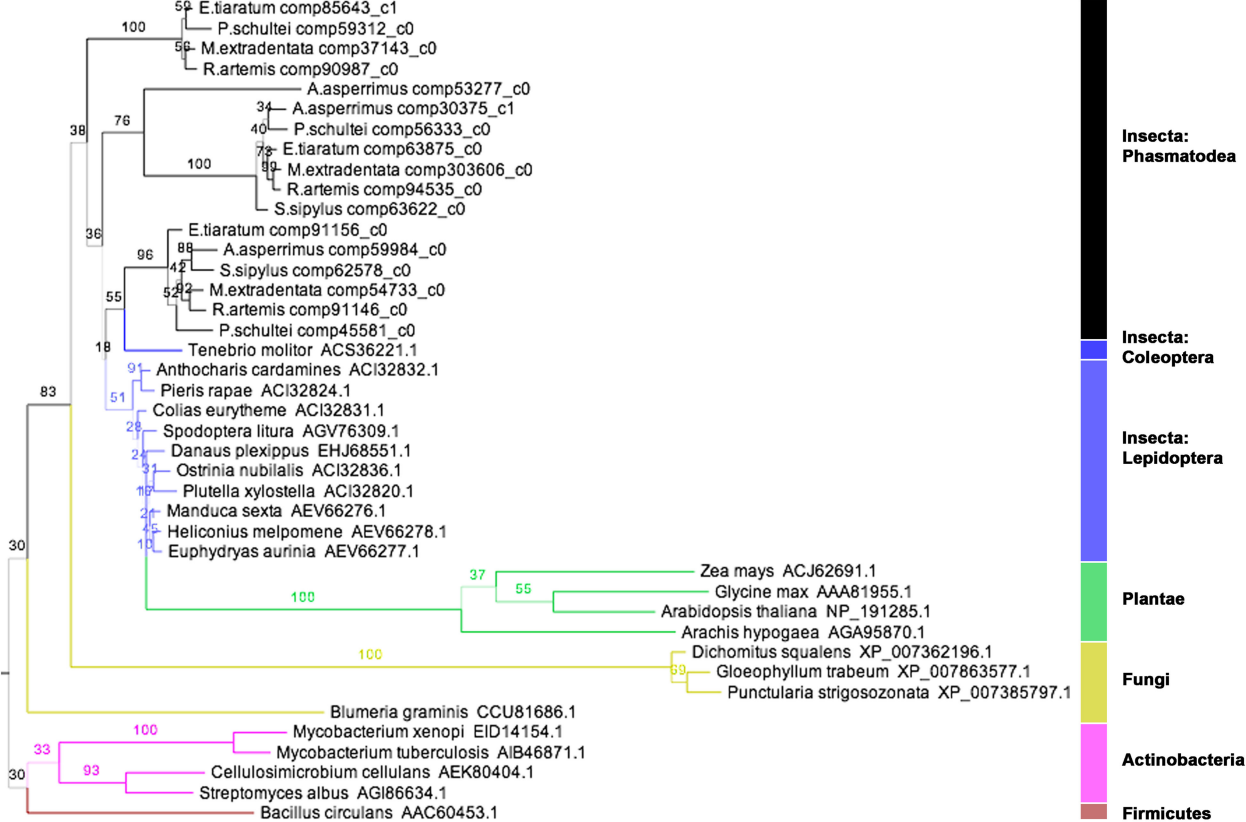

E.tiaratum comp85643_c1
P.schultei comp59312_c0
M.extradentata comp37143_c0
R.artemis comp90987_c0
A.asperrimus comp53277_c0
A.asperrimus comp30375_c1
P.schultei comp56333_c0
E.tiaratum comp63875_c0
M.extradentata comp303606_c0
R.artemis comp94535_c0
S.sipylus comp63622_c0
E.tiaratum comp91156_c0
A.asperrimus comp59984_c0
S.sipylus comp62578_c0
M.extradentata comp54733_c0
R.artemis comp91146_c0
P.schultei comp45581_c0
Tenebrio molitor ACS36221.1
Anthocharis cardamines ACI32832.1
Pieris rapae ACI32824.1
Colias eurytheme ACI32831.1
Spodoptera litura AGV76309.1
Danaus plexippus EHJ68551.1
Ostrinia nubilalis ACI32836.1
Plutella xylostella ACI32820.1
Manduca sexta AEV66276.1
Heliconius melpomene AEV66278.1
Euphydryas aurinia AEV66277.1
Zea mays ACJ62691.1
Glycine max AAA81955.1
Arabidopsis thaliana NP_191285.1
Arachis hypogaea AGA95870.1
Dichomitus squalens XP_007362196.1
Gloeophyllum trabeum XP_007863577.1
Punctularia strigosozonata XP_007385797.1
Blumeria graminis CCU81686.1
Mycobacterium xenopi EID14154.1
Mycobacterium tuberculosis AlB46871.1
Cellulosimicrobium cellulans AEK80404.1
Streptomyces albus AGI86634.1
Bacillus circulans AAC60453.1
Insecta: Phasmatodea
Insecta: Coleoptera
Insecta: Lepidoptera
Plantae
Fungi
Actinobacteria
Firmicutes
0.7

Figure 9

Additional files provided with this submission:

Additional file 1: 1531942522138963_add1.xls, 27K
http://www.biomedcentral.com/imedia/1131788598146781/supp1.xls
Additional file 2: 1531942522138963_add2.xls, 29K
http://www.biomedcentral.com/imedia/5870135014678169/supp2.xls
Additional file 3: 1531942522138963_add3.xls, 25K
http://www.biomedcentral.com/imedia/5398515051467816/supp3.xls
Additional file 4: 1531942522138963_add4.xls, 42K
http://www.biomedcentral.com/imedia/9307854191467816/supp4.xls
Additional file 5: 1531942522138963_add5.xls, 24K
http://www.biomedcentral.com/imedia/1953606342146781/supp5.xls
Additional file 6: 1531942522138963_add6.xls, 11K
http://www.biomedcentral.com/imedia/5079358081467817/supp6.xls
Additional file 7: 1531942522138963_add7.xls, 22K
http://www.biomedcentral.com/imedia/1678372074146781/supp7.xls
Additional file 8: 1531942522138963_add8.xls, 60K
http://www.biomedcentral.com/imedia/1715491221467817/supp8.xls
Additional file 9: 1531942522138963_add9.txt, 0K
http://www.biomedcentral.com/imedia/2465331111467817/supp9.txt
Additional file 10: 1531942522138963_add10.pdf, 3K
http://www.biomedcentral.com/imedia/1935738471467817/supp10.pdf
Additional file 11: 1531942522138963_add11.xls, 3189K
http://www.biomedcentral.com/imedia/1029747732146781/supp11.xls
Additional file 12: 1531942522138963_add12.xls, 3182K
http://www.biomedcentral.com/imedia/2862394814678170/supp12.xls
Additional file 13: 1531942522138963_add13.xls, 2764K
http://www.biomedcentral.com/imedia/8439712231467817/supp13.xls
Additional file 14: 1531942522138963_add14.xls, 2794K
http://www.biomedcentral.com/imedia/3763058691467817/supp14.xls