Pyrosequencing of the midgut transcriptome of the poplar leaf beetle Chrysomela tremulae reveals new gene families in Coleoptera

Yannick Pauchet, Paul Wilkinson, Manuella van Munster, Sylvie Augustin, David Pauron, Richard Ffrench-Constant

To cite this version:

Yannick Pauchet, Paul Wilkinson, Manuella van Munster, Sylvie Augustin, David Pauron, et al.. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle Chrysomela tremulae reveals new gene families in Coleoptera. Insect Biochemistry and Molecular Biology, 2009, 39 (5-6), pp.403-413. 10.1016/j.ibmb.2009.04.001 . hal-02668241

HAL Id: hal-02668241
https://hal.inrae.fr/hal-02668241v1

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

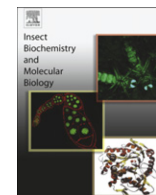
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Insect Biochemistry and Molecular Biology xxx (2009) 1–11

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb



Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera

Yannick Pauchet [a,*], Paul Wilkinson [a], Manuella van Munster [b], Sylvie Augustin [c], David Pauron [b], Richard H. ffrench-Constant [a]

[a] School of Biosciences, University of Exeter in Cornwall, Tremough Campus, Penryn TR10 9EZ, UK
[b] INRA, UMR 1301 Interactions Biotiques et Santé Végétale, 06903 Sophia Antipolis cedex, France
[c] INRA, UR 633 Zoologie Forestière, F-45000 Orléans, France

ARTICLE INFO

Article history:
Received 10 February 2009
Received in revised form
2 April 2009
Accepted 2 April 2009

Keywords:
Insect midgut
Chitin deacetylase
Cellulase
Endopolygalacturonase
Chrysomela
Bt resistance

ABSTRACT

The insect midgut is the primary target site for Bt-derived insecticides and Bt alternatives. However, despite extensive recent study, the precise role and nature of different Bt receptors remains a subject of considerable debate. This problem is fuelled by a lack of understanding of the genes expressed in the insect midgut and their physiological roles. The poplar leaf beetle, *Chrysomela tremulae*, is an important model for understanding the mode of action of, and resistance to, coleopteran-specific Bt toxins and currently shows the only known naturally occurring case of resistance to Cry3A toxins. Moreover it belongs to the same family as the corn rootworm, *Diabrotica virgifera*, an economically important beetle pest and target of hybrid corn expressing Cry3 toxins. Pyrosequencing is a fast and efficient way of defining the transcriptome of specific insect tissues such as the larval midgut. Here we use 454 based pyrosequencing to sample the larval midgut transcriptome of *C. tremulae*. We identify candidate genes of putative Bt receptors including transcripts encoding cadherin-like proteins, aminopeptidase N and alkaline phosphatase. We also describe a wealth of new transcripts predicting rapidly evolving gene families involved in plant tissue digestion, which have no homologs in the genome of the stored product pest the Red Flour beetle, *Tribolium castaneum*.

© 2009 Published by Elsevier Ltd.

UNCORRECTED PROOF

1. Introduction

Traditionally the major target site for small molecule insecticides has been the insect nervous system (Buckingham et al., 2005; Dong, 2007; ffrench-Constant et al., 2004; Millar and Denholm, 2007; Raymond-Delpech et al., 2005). However with the advent of Bt toxins, and Bt transformed insect resistant crops, the midgut has become the primary focus for the development of novel insecticidal proteins (Pigott and Ellar, 2007). Interest in insect midgut physiology has been strengthened by concerns over emerging Bt resistance and by ongoing research into a growing number of candidate Bt receptors and how these may become altered in Bt resistant insects (Pigott and Ellar, 2007). Identifying and understanding midgut receptors will also be important in the study of new toxins proposed as alternatives to toxins from *Bacillus*, such as the Toxin complexes (Tc's) from *Photorhabdus* and *Xenorhabdus* bacteria (Bowen et al., 1998; ffrench-Constant et al., 2007). However, despite extensive recent study, the precise role and nature of different Bt receptors remains a subject of considerable debate (Pigott and Ellar, 2007; Soberon et al., 2009). This problem is fuelled by a lack of understanding of the genes expressed in the insect midgut and their respective roles in digestion, response to invading micro-organisms, homeostatic regulation, and the production and remodeling of the peritrophic membrane. We aim to address this problem by describing the midgut transcriptome of representatives of key insect pest groups such as leaf feeding beetles and Lepidoptera.

Beetles represent approximately half of all insect species recorded to date and within this group the chrysomelids are one of the largest families (Hunt et al., 2007), many of which are important defoliating and/or root damaging pests such as the corn rootworm *Diabrotica virgifera* (Gray et al., 2009). With the recent sequencing of the complete genome of the Red Flour Beetle, *Tribolium castaneum*, we are now uniquely poised to understand more about the genomics and transcriptomics of beetles. Moreover, given the rather unusual lifestyle of *T. castaneum* in feeding exclusively on stored products, comparisons with the commoner lifestyle of leaf-, root- or bark-feeding are a priority. The poplar leaf beetle, *Chrysomela*

Abbreviations: EST, expressed sequence tag; PM, peritrophic membrane; Bt, *Bacillus thuringiensis* toxin; SNP, single nucleotide polymorphism.

* Corresponding author. Tel.: +44 1326 371 852; fax: +44 1326 253 638.
E-mail address: y.pauchet@exeter.ac.uk (Y. Pauchet).

0965-1748/$ – see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.ibmb.2009.04.001

tremulae, is an important model for understanding Bt receptors and Bt resistance, and this is currently the only known beetle for which Cry3A resistance alleles have been found in the field (Genissel et al., 2003a). A Cry3A resistant *C. tremulae* strain was selected on Bt transformed poplar trees expressing the Cry3A toxin (Genissel et al., 2003b). This strain was derived from an isofemale line established from field-caught insects that generated F2 offspring that survived on this Bt poplar clone (Augustin et al., 2004). Resistance to Cry3A in *C. tremulae* is under control of a single, completely recessive, autosomal trait (Augustin et al., 2004), suggesting that changes in a single receptor, or other gene product, may be involved in resistance. Resistance to Cry3A in the related *Chrysomela scripta* can be overcome using Cyt1Aa (Federici and Bauer, 1998), a protein from a different class of Bt toxins, suggesting some potential specificity in the nature of the resistance mechanism. In the Lepidoptera, Bt resistance has been associated with changes in midgut associated aminopeptidase N, cadherin-like proteins, alkaline phosphatase and, via work in model nematodes, midgut glycolipids (Pigott and Ellar, 2007). However the molecular basis of Cry3A resistance in beetles has not been reported to date.

Pyrosequencing is a fast and efficient way of defining the transcriptome of specific insect tissues or cells and has recently been used with great effect to define the transcriptome of immune related tissues in *Manduca sexta* larvae (Zou et al., 2008). 454 based sequencing of Expressed Sequence Tags (ESTs) greatly increases the number and depth of sequence contigs that can be achieved using capillary based sequencing. Further, normalization of cDNA pools prior to sequencing reduces the frequency of highly abundant transcripts and ensures that a more even representation of transcripts is sampled during sequencing. Thus 454 sequencing looks poised to geometrically increase the number of known transcripts in many insect tissues. Here we use 454 based pyrosequencing to sample the larval midgut transcriptome of *C. tremulae*. We identify several classes of putative Bt receptors such as transcripts encoding putative cadherin-like proteins, aminopeptidase N and alkaline phosphatase. We also describe a wealth of putative new gene products which are part of rapidly evolving gene families involved in plant cell wall digestion. These new gene families are not present in the genome of *T. castaneum* but are common in EST collection from other plant feeding beetles, suggesting that additional leaf feeding genomic models, such as *C. tremulae*, will be needed in order to fully understand the physiology of the Coleoptera.

2. Methods and materials

2.1. Insect rearing and midgut RNA preparation

C. tremulae larvae for midgut dissection were obtained from a strain established from a single mated pair collected in Vatan, France, in 1999 (Augustin et al., 2004). The strain (Vatan125), originated from the offspring of an isofemale line lacking alleles conferring resistance to the Cry3Aa toxin, was maintained in standard rearing conditions, in a growth chamber at 20 °C with a photoperiod of 16:8 (L:D) at URZF-INRA Orléans. Larvae and adults were reared on fresh mature leaves detached from greenhouse-grown poplar hybrid clone (*Populus tremula* × *Populus tremuloides*, INRA #353-38). Three day old third-instar larvae were used for dissection.

For RNA isolation, midguts were dissected from 10 larvae into PBS buffer. Midguts were cleared of food, peritrophic membrane and Malpighian tubules, before being flash frozen in liquid nitrogen and stored at −80 °C prior to use. RNA was subsequently isolated using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. Genomic DNA contamination was removed by DNAse treatment (TURBO DNAse, Ambion) for 30 min at 37 °C. Midgut RNA was further purified by using the RNeasy MinElute Clean up Kit (Qiagen) following the manufacturer's protocol and eluted in 20 µl of RNA storage solution (Ambion).

2.2. Midgut cDNA library preparation

Full-length, enriched, cDNAs were generated from 2 µg of total RNA using the SMART PCR cDNA synthesis kit (BD Clontech) following the manufacturer's protocol. Reverse transcription was performed with the PrimeScript reverse transcription enzyme (Takara) for 60 min at 42 °C and 90 min at 50 °C. In order to prevent over-representation of the most common transcripts, the resulting double-stranded cDNAs were normalized using the Kamchatka crab duplex-specific nuclease method (Trimmer cDNA normalization kit, Evrogen) (Zhulidov et al., 2004). The resulting normalized cDNA midgut library was used for 454 pyrosequencing (Margulies et al., 2005).

2.3. Pyrosequencing, sequence preprocessing and assembly

For 454 pyrosequencing, a cDNA aliquot was sent to the Advanced Genomics facility at the University of Liverpool (http://www.liv.ac.uk/agf). A single full plate run was performed on the 454 GS-FLX pyrosequencer (Roche Applied Science) using 3 µg of normalized cDNAs processed by the "shotgun" method. The reads obtained were preprocessed by removing PolyA tails and SMART adapters using custom written Perl scripts kindly provided by Mark Blaxter of the University of Edinburgh. The processed reads were then clustered using the MIRA v2.9.26x3 assembler with the "*de novo*, normal, EST, 454" parameters. Specifying a minimum read length of 40 nt, a minimum sequence overlap of 40 nt and a minimum percentage overlap identity of 80%.

2.4. Blast homology searches and sequence annotation

Homology searches (BLASTx and BLASTn) of unique sequences and functional annotation by gene ontology terms (GO; www.geneontology.org), InterPro terms (InterProScan, EBI), enzyme classification codes (EC), and metabolic pathways (KEGG, Kyoto Encyclopedia of Genes and Genomes) were determined using the BLAST2GO software suite v2.3.1 (www.blast2go.de) (Conesa and Götz, 2008). Homology searches were performed remotely on the NCBI server through QBLAST, and followed a sequential strategy. First, sequences were searched against an NCBI non-redundant (nr) protein database via BLASTx using an *E*-value cut-off of 10^{-3} and selecting predicted polypeptides of a minimum length of 10 amino acids. Second, sequences for which no BLASTx hit were searched again by BLASTn, against an NCBI nr nucleotide database using an *E*-value cut-off of 10^{-10}.

For gene ontology mapping, the program extracts the GO terms associated with homologies identified with NCBI's QBLAST and returns a list of GO annotations represented as hierarchical categories of increasing specificity. BLAST2GO allows the selection of a significance level for the False Discovery Rate (FDR) which was used as a cut-off at the 0.05% probability level. Then, GO terms were modulated using the annotation augmentation tool ANNEX (Myhre et al., 2006) followed by GOSlim. GOSlim consists of a subset of the gene ontology vocabulary encompassing key ontological terms and a mapping function between the full GO and the GOSlim. Here, we used the "generic" GOSlim mapping (goslim_generic.obo) available in BLAST2GO. The data presented represent the level 2 analysis, illustrating general functional categories. Enzyme classification codes and KEGG metabolic pathway annotations are generated from the direct mapping of GO terms to their enzyme code equivalents. Finally, InterPro searches were performed remotely from

BLAST2GO to the InterProEBI web server. The default settings of BLAST2GO were used in every annotation step. For more details on BLAST2GO functionalities see Conesa and Götz (2008).

2.5. Capillary based sequencing of highly abundant messages

The over-abundant ~1.3 kb band observed in the non-normalized cDNA lane (Fig. 1) was gel-extracted and purified using the Illustra GFX PCR DNA and Gel Band Purification kit (GE healthcare). The resulting DNA was re-amplified by 25 PCR cycles using the 5′ PCR Primer II A provided in the SMART PCR cDNA synthesis kit (BD Clontech). After checking on a 1.5% agarose gel that only one band of ~1300 bp was re-amplified, the PCR product was cloned into the pCR4 TOPO vector (Invitrogen) and ligations transformed into chemically competent *Escherichia coli* TOP10 (Invitrogen). For capillary sequencing, DNA template was prepared on a Nucleoplex automated system using the Nucleoplex Plasmid Mini Prep kit (Tepnel Life Sciences). Capillary sequencing was carried out on an ABI 3130xl automatic DNA sequencer (Applied Biosystems). Unique cDNAs found in at least two clones were fully sequenced by primer walking. The unique cDNA sequences obtained were searched by BLAST and fully annotated using BLAST2GO. Sequences were submitted to Genbank with accession numbers FJ654706–FJ654727.

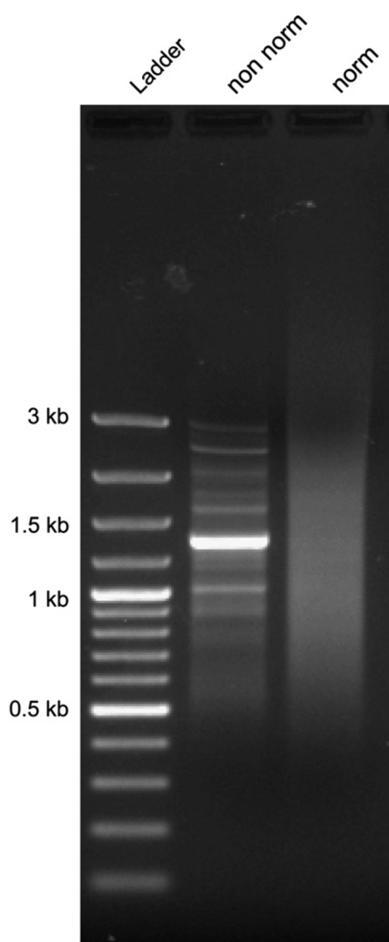


Fig. 1. Agarose gel of *C. tremulae* midgut cDNA before (non-norm) and after (norm) normalization. Note the high relative abundance of several discrete bands in the non-normalized sample which disappear following normalization. The over-abundant transcripts (~1.3 kb) were excised from the gel, cloned and sequenced via capillary sequencing to characterize highly abundant larval midgut transcripts (see text). Size markers are shown at left.

2.6. Phylogenetic analysis

MAFFT software (Multiple Alignment using Fast Fourier Transform, www.ebi.ac.uk/mafft/) was used to perform multiple sequence alignment of the chitin deacetylase catalytic domains prior to subsequent phylogenetic analysis. Predicted protein sequences corresponding to cellulases, deduced from *C. tremulae* 454 contigs and coleopteran-specific ESTs collected from Genbank db_est database, were screened for the presence of a putative signal peptide using the SignalP software. After removal of the predicted signal peptide, sequences were aligned using MAFFT software prior to Phylogenetic analysis. The MEGA4.0 software (Tamura et al., 2007) was used to construct the consensus phylogenetic tree using the UPGMA method (Unweighted Pair Method with Arithmetic Mean). To evaluate the branch strength of the phylogenetic tree, bootstrap analysis of 1000 replications was performed.

3. Results

3.1. Pyrosequencing, assembly and annotation

Normalization of the midgut cDNA resulted in reduction of the over-abundant 0.8–3.0 kb transcripts and production of an even distribution of transcripts ranging from 0.5 to 3.0 kb in size (Fig. 1). 454-mediated pyrosequencing of the normalized library, using a full sequencing plate of a Roche GS-FLX sequencer, generated a total of 264,698 reads of raw nucleotide sequence data with an average read length of 226 bp (Table 1). After quality scoring using MIRA v2.9.26x3, 210,885 of these reads were taken forward for assembly. Assembly resulted in 23,238 contigs and a residual 87 singletons, defined as sequences that did not assemble into a contig using the defined assembly parameters. The mean contig size was 418 bp, with contigs ranging from as small as 40 bp to as large as 2959 bp (Table 1).

For BLAST annotation, contigs were first searched using BLASTx against the non-redundant NCBI protein database using a cut-off *E*-value of 10^{-3}. Contigs returning below cut-off BLASTx results were then re-searched using BLASTn against the non-redundant NCBI nucleotide database using a cut-off *E*-value of 10^{-10}, using this sequential approach, just over half (56.1%) of all contigs returned an above cut-off BLAST result (Table 1).

3.2. Functional classifications of predicted proteins

The *E*-value and similarity distributions of the *C. tremulae* BLAST hits against the non-redundant database are shown in Fig. 2. Most (67%) of the BLAST hits are to the *T. castaneum* genome (Fig. 2C), as to date this is the only fully sequenced beetle genome and therefore *T. castaneum* sequences currently represent the vast majority of beetle sequences in current databases. Gene ontology (GO) assignments

Table 1
Summary statistics for *C. tremulae* midgut EST assembly and annotation. Q2

Assembly	
Total number of reads	264,698
Average read length	226 bp
Number of reads that entered in the assembly[1]	210,885
Total number of contigs	23,238
Total number of singlets	87
Average contig size (range)	418 bp (40–2959 bp)
Annotation	
% contigs with a blast hit against nr[2]	56.08%
% contigs with at least 1 GO term	28.73%
% contigs with an EC number	6.65%
% contigs with at least one IPR	26.81%

Please cite this article in press as: Pauchet, Y., et al., Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela...*, *Insect Biochemistry and Molecular Biology (2009)*, doi:10.1016/j.ibmb.2009.04.001

and enzyme classifications (EC) were used to classify the functions of the predicted midgut proteins (Fig. 3). A dominance of metabolic, catalytic and binding activities is predicted (Fig. 3B and C), as expected for midgut proteins. Within the range of enzyme activities hydrolases dominate accounting for over half (51%) of midgut enzymes (Fig. 3D). The InterPro database was also used to classify the likely functions of predicted midgut proteins and a summary of the 20 most frequent classifications is shown in Table 2. Of these, enzymes involved in protein digestion (including a range of cysteine and serine proteinases) or xenobiotic metabolism (cytochrome P450s, glutathione-S-transferases and carboxylesterases) are dominant, as again predicted for the role of the insect gut in digestion and xenobiotic metabolism. In Table 3, we list candidate genes of interest for different functional roles discussed below, including cell wall degradation, putative Bt receptors and peritrophic membrane associated proteins. Six contigs predict a *C. tremulae* cadherin-like protein showing 40–64% amino acid identity to other coleopteran and lepidopteran cadherin-like proteins. Contig 2921 shows 64% identity and 80% similarity to the cadherin-like protein from *D. virgifera* (Sayed et al., 2007) and is therefore a candidate Bt receptor. An alignment of contig 2921 with coleopteran cadherin-like proteins is shown in Fig. 4. The predicted protein sequence of contig 2921 corresponds to part of the transmembrane spanning repeats and to the whole cytosolic region of cadherin-like proteins.

3.3. Identification of microsatellite repeats

Microsatellite loci with ≥5 di-, tri-, tetra- and pentanucleotide repeat units were identified from 67 *C. tremulae* 454 contigs. Thanks to the high coverage enabled by 454 pyrosequencing, variation in the number of tandem repeats could be assessed directly from the assembly of 3 contigs (Fig. 5). These 3 loci were found in the 5′-UTR of a transcript similar to a *T. castaneum* sodium-dependent phosphate transporter XP_974340.1 (Fig. 5A), the 3′-UTR of a homolog of a *Nasonia vitripennis* predicted protein XP_001607853.1 (Fig. 5B), and the CDS of a homolog of a *Bombyx mori* fumarylacetoacetate hydrolase NP_001103763.1, matching codons for Tyr (TAT) and Ile (ATA) amino acids (Fig. 5C).

3.4. Capillary sequencing of abundant midgut transcripts

Following the excision of abundant 1.2–1.5 kb transcripts from the agarose gel of non-normalized midgut cDNA, the resulting product was cloned and individual cDNAs sequenced via conventional capillary sequencing. A representative sample of the most abundant cDNAs recovered is listed in Table 4, whilst a complete list of transcripts and their associated InterPro predictions are given in the Supplementary materials (Table S1 and S2). The most abundant transcripts, with 40 clones recovered, were two isoforms of a chitin

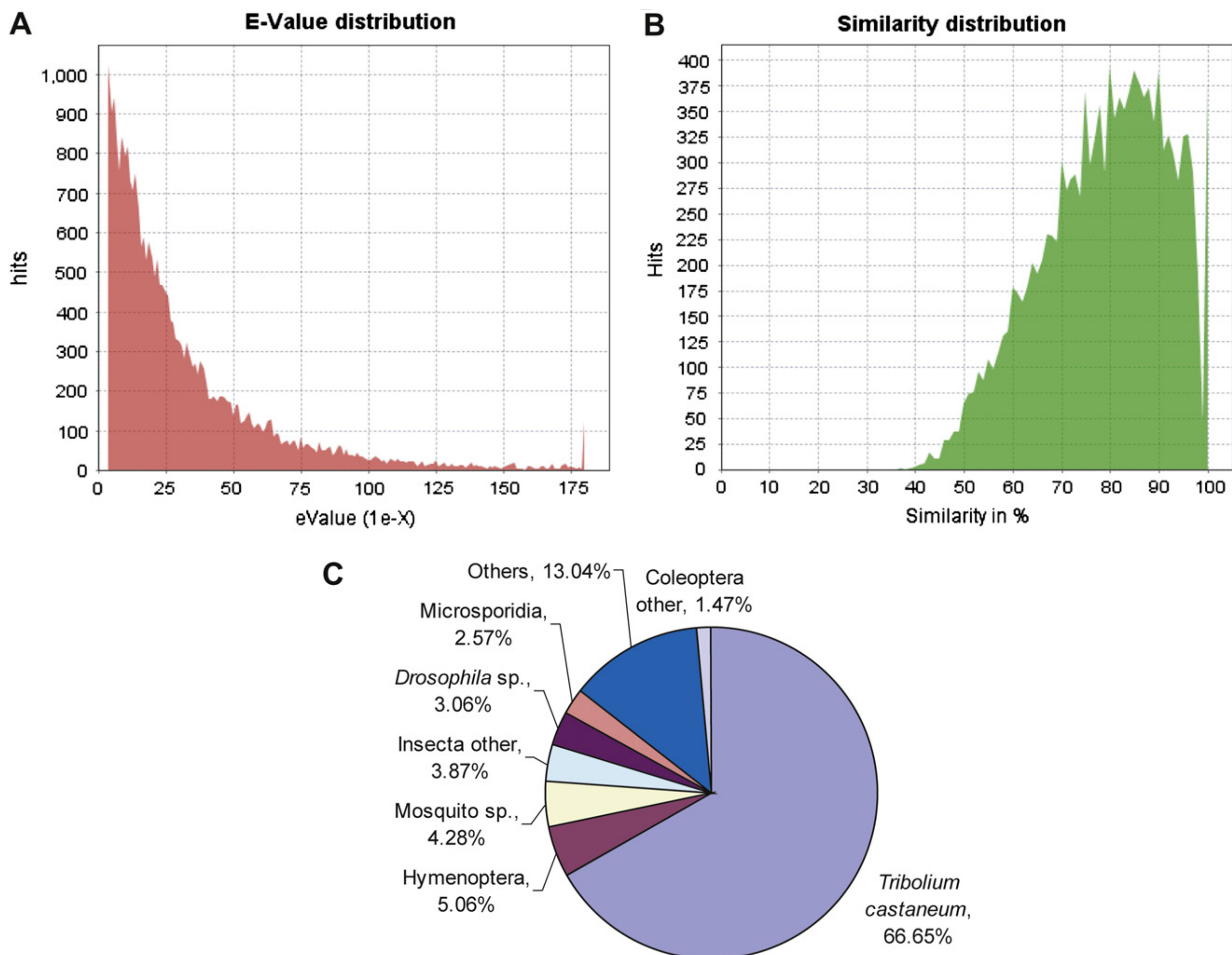


Fig. 2. Summary of homology searches (BLASTx and BLASTn) of *C. tremulae* larval midgut 454 data against the non-redundant database at NCBI. (A) *E*-value distribution of the top BLAST hit for each unique sequence. Cut-offs used in each case were 10^{-3} for BLASTx and 10^{-10} for BLASTn. (B) Similarity distribution of the top BLAST hit for each unique sequence. (C) Species distribution of the top BLAST hit for each unique sequence. Note that nearly 67% of top hits are to the beetle *Tribolium* whose complete genome has recently been sequenced.

Please cite this article in press as: Pauchet, Y., et al., Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela*..., *Insect Biochemistry and Molecular Biology (2009), doi:10.1016/j.ibmb.2009.04.001*

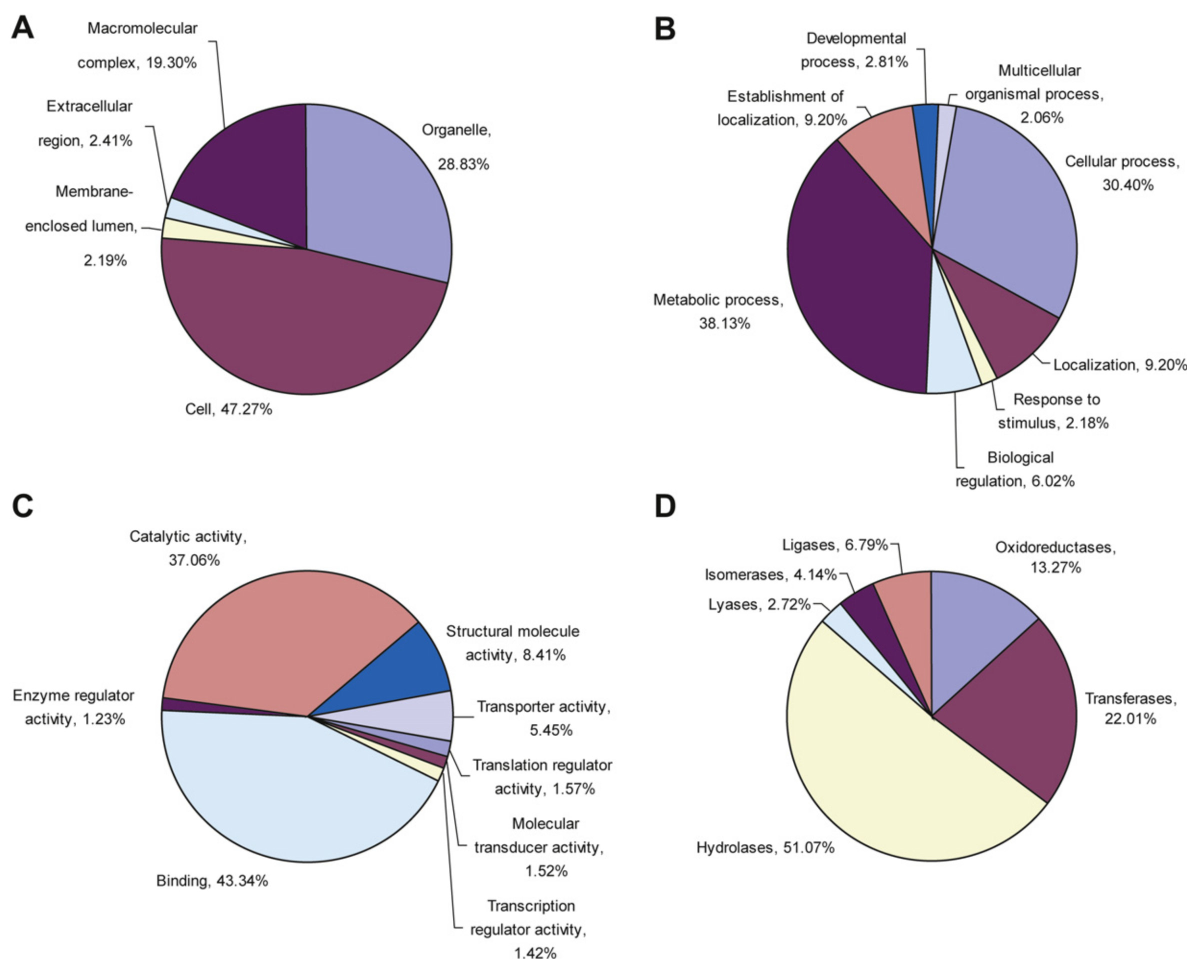


Fig. 3. Gene Ontology (GO) assignments and Enzyme Classifications (EC) for the *C. tremulae* larval midgut transcriptome. (A) Cellular component GO terms at level 2. (B) Biological process GO terms at level 2. (C) Molecular function GO terms at level 2. (C) General EC terms: Oxidoreductases (EC:1.x.x), Transferases (EC:2.x.x), Hydrolases (EC: 3.x.x), Lyases (EC:4.x.x), Isomerases (EC:5.x.x), and Ligases (EC:6.x.x). Note that one sequence can be associated with more than one GO term (see text for discussion).

Table 2
Summary of TOP20 of InterPro families and domains represented in the *C. tremulae* larval midgut transcriptome.

Family			Domain		
InterPro	Frequency	Description	InterPro	Frequency	Description
IPR013128	51	Peptidase C1A, papain	IPR016040	66	NAD(P)-binding
IPR002198	39	Short-chain dehydrogenase/reductase	IPR002018	52	Carboxylesterase, type B
IPR013753	33	Ras	IPR000668	45	Peptidase C1A, papain C-terminal
IPR001806	30	Ras GTPase	IPR017853	44	Glycoside hydrolase, catalytic core
IPR001128	25	Cytochrome P450	IPR002048	41	Calcium-binding EF-hand
IPR001360	24	Glycoside hydrolase, family 1	IPR011009	41	Protein kinase-like
IPR001993	24	Mitochondrial substrate carrier	IPR000719	38	Protein kinase, core
IPR002347	24	Glucose/ribitol dehydrogenase	IPR009003	37	Peptidase, trypsin-like serine and cysteine
IPR001353	23	20S proteasome	IPR001254	35	Peptidase S1/S6, chymotrypsin/Hap
IPR001395	21	Aldo/keto reductase	IPR004046	34	Glutathione-S-transferase, C-term
IPR004119	21	Protein of unknown function DUF227	IPR007087	33	Zinc finger, C2H2-type
IPR004000	19	Actin/actin-like	IPR011046	33	WD40 repeat-like
IPR015609	18	Heat shock protein HSP40	IPR016196	31	MFS general substrate transporter
IPR015643	18	Peptidase C1A, cathepsin B	IPR015880	29	Zinc finger, C2H2-like
IPR000734	17	Lipase	IPR009072	25	Histone-fold
IPR007114	17	Major facilitator family	IPR000504	24	RNA recognition motif, RNP-1
IPR006649	16	Like-Sm ribonucleoprotein	IPR002557	24	Chitin-binding, peritrophin A
IPR000743	15	Glycoside hydrolase, family 28	IPR016135	22	Ubiquitin-conjugating enzyme/RWD-like
IPR001316	15	Peptidase S1A, streptogrisin	IPR017442	22	Serine/threonine protein kinase-related
IPR001930	15	Peptidase M1, membrane alanine aminopeptidase	IPR016024	21	Armadillo-type fold

Please cite this article in press as: Pauchet, Y., et al., Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela*..., *Insect Biochemistry and Molecular Biology (2009), doi:10.1016/j.ibmb.2009.04.001*

Table 3
Selection of genes of interest related to the *C. tremulae* larval midgut physiological functions.

	EC number	Total number of contigs
Plant/fungi cell wall degradation		
Cellulase	3.2.1.4	2
Cellulose 1,4-β-cellobiosidase	3.2.1.91	6
Endopolygalacturonase (pectinase)	3.2.1.15	37
Endo-β-1,3-glucanase	3.2.1.39	2
Bt Cry3 toxin putative binding partners		
Cadherin-like protein	–	6
Aminopeptidase N	3.4.11.2	11
Alkaline phosphatase	3.1.3.1	1
Peritrophic membrane biosynthesis, degradation and remodeling		
Chitinase	3.2.1.14	36
Chitin synthase	2.4.1.16	7
Chitin deacetylase	3.1.5.41	13
Peritrophin (mucin)	–	7
General digestion		
Cysteine proteinase all types	–	129
Serine proteinase all types	–	69
Aminopeptidase all types	–	33
Carboxypeptidase all types	–	25
Dipeptidylpeptidase all types	–	3
Lipase	3.1.1.3	17
α-amylase	3.2.1.1	3
α-glucosidase (maltase)	3.2.1.20	10
β-glucosidase	3.2.1.21	22

deacetylase differing by eight SNPs resulting in only two amino acid changes. These two isoforms are putatively two alleles of the same chitin deacetylase gene. Phylogenetic comparison with chitin deacetylases from a range of different insects (Fig. 6), shows that these belong to the midgut-specific Group V, according to the classification described by Dixit et al. (2008), with the closest relative being chitin deacetylase 6 from *T. castaneum*. A full amino acid alignment of the putative catalytic domains used to derive the phylogeny is shown in the Supplementary materials (Fig. S1). A total of 13 contigs, corresponding to the two alleles described above, and also to homologs of *T. castaneum* chitin deacetylase 2 and 5 were found in the 454 data (Table 3).

3.5. Identification of new beetle gene families

After searching *C. tremulae* contigs using BLASTx against the *T. castaneum* gene objects (16,404 entries) (Richards et al., 2008), 12,897 positive hits were returned corresponding to 5489 (33.5%) unique *T. castaneum* gene objects. Contigs returning below cut-off BLASTx results were re-searched against a non-redundant arthropod database returning a further 428 positive hits. A large portion of these contigs that did not hit any *Tribolium* predicted gene objects were identified as putative plant tissue degrading enzymes such as cellulases (EC:3.2.1.4) and endopolygalacturonases (EC:3.2.1.15).

To determine if such genes are common in other Coleoptera, although absent from *T. castaneum*, we used the predicted proteins

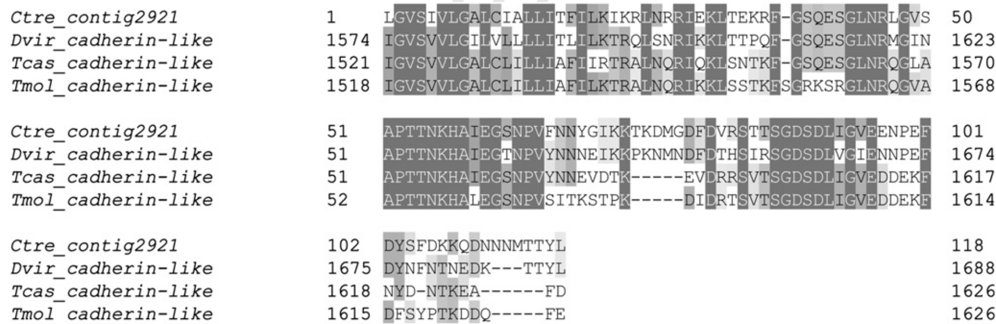


Fig. 4. Amino acid alignment of known coleopteran cadherin-like genes with *C. tremulae* contig 2921, showing part of the predicted transmembrane spanning repeats. Sequences were aligned using the MAFFT multiple alignment program and identical residues are boxed with dark shading, and conserved residues boxed with light shading. Sequences are from *Diabrotica virgifera* (Dvir, AAV88529.2), *Tribolium castaneum* (Tcas, XP_971388.2), *Tenebrio molitor* (Tmol, ABL86001.1) and *C. tremulae* (Ctre, contig_2921).

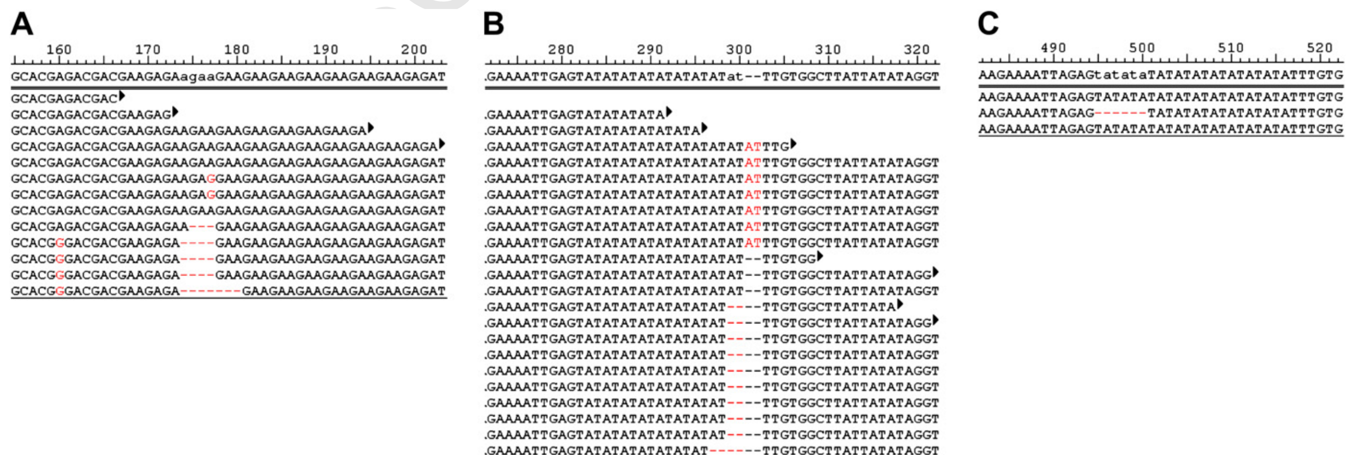


Fig. 5. Polymorphic microsatellite repeats found in the *C. tremulae* 454 dataset. (A) $(GAA)_n$ found in the 5′-UTR of a putative sodium-dependent phosphate transporter transcript. (B) $(TA)_n$ from the 3′-UTR of a predicted protein transcript. (C) $(TA)_n$ from the CDS of a putative fumarylacetoacetate hydrolase transcript.

Please cite this article in press as: Pauchet, Y., et al., Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela*..., Insect Biochemistry and Molecular Biology (2009), doi:10.1016/j.ibmb.2009.04.001

Table 4
Capillary sequencing of over-abundant transcripts (~1.3 kb) in the non-normalized midgut cDNA.

# clones	% total	Name	cDNA (bp)	Protein (aa)	Blastp vs nr	Species	*E*-Value
40	41.7	Chitin deacetylase	1260	376	Chitin deacetylase 6	*T. castaneum*	3e−102
10	10.42	Aspartic proteinase-1	1245	386	Cathepsin D	*T. castaneum*	3e−134
10	10.42	16S rRNA	1306	–	–	–	–
3	3.13	Chitinase-1	1278	389	Chitinase	*P. cochleariae*	1e−118
3	3.13	Cysteine proteinase-1	1163	323	Cathepsin L-like	*P. cochleariae*	4e−138
3	3.13	Endopolygalacturonase-1	1205	370	Polygalacturonase	*S. oryzae*	5e−97
2	2.1	Tetraspanin-1	1136	305	Tetraspanin 29fb	*T. castaneum*	3e−67
2	2.1	Unknown protein-1	1275	397	Hypothetical protein	*T. castaneum*	8e−69

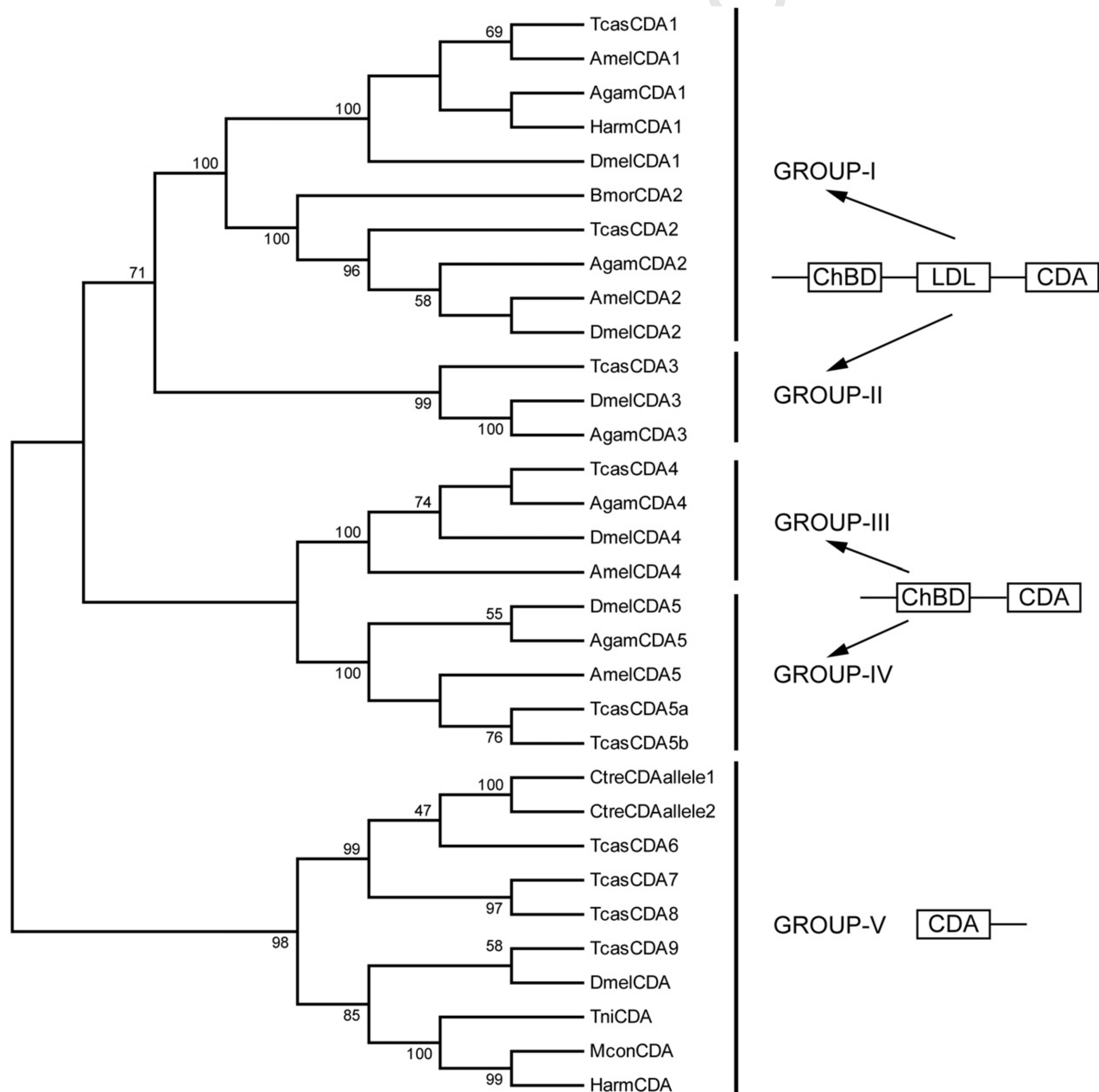


Fig. 6. Phylogenetic analysis of putative chitin deacetylases from different insects. Phylogenetic analysis was conducted in MEGA4.0 using the UPGMA method and the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). ChBD: Chitin-binding domain; LDL: Low density lipoprotein domain; CDA: Chitin deacetylase domain. *Tribolium castaneum* (Tcas), *Apis mellifera* (Amel), *Anopheles gambiae* (Agam), *Helicoverpa armigera* (Harm), *Bombyx mori* (Bmor), *Drosophila melanogaster* (Dmel), *Trichoplusia ni* (Tni), *Mamestra configurata* (Mcon), *C. tremulae* (Ctre). Note that the alignment of the CDA domains used in this phylogeny is provided as a Supplementary material (Fig. S1).

of the corresponding *C. tremulae* contigs to search public coleopteran EST collections present in the Genbank db_est database. We retrieved 130 ESTs corresponding to cellulases that clustered into 20 unique contigs, and 29 ESTs corresponding to pectinases that clustered into 18 unique contigs. A further three coleopteran cellulase sequences were retrieved from Genbank together with two endopolygalacturonase sequences. All these genes were found in coleopteran species feeding on fresh plant tissues such as the pine engraver *Ips pini*, the western corn rootworm *D. virgifera*, the Colorado potato beetle *Leptinotarsa decemlineata*, the coffee berry borer *Hypothenemus hampei*, the mustard beetle *Phaedon cochleariae* and the citrus root weevil *Diaprepes abbreviatus*. These sequence data are summarized in Table S3.

The SignalP software predicted the presence of a putative signal peptide in all the cellulase protein sequences that appeared to be full-length CDS. An alignment of all these sequences is provided as Supplementary material (Fig. S2). In addition, an InterProScan search predicted a glycoside hydrolase family 45 (GH45) conserved domain (IPR000334) in every sequence. The active site consensus pattern for GH45 ([STA]-T-R-Y-[FYW]-D-x(5)-[CA]) is also conserved in all sequences (Fig. 7A). As many as six different cellulase sequences could be found in some species such as *I. pini*, but only two contigs were present in the *C. tremulae* 454 data (Fig. 7A). A phylogenetic analysis suggested the presence of at least five sub-classes in this gene family (Fig. 7B).

Thirty-seven contigs corresponding to putative endopolygalacturonase genes were found in the *C. tremulae* dataset suggesting a highly diverse gene family. An alignment of the predicted proteins deduced from the 11 longest contigs together with the sequence found in the abundant 1.3 kb band (Pect-1, Table 4) is shown in Fig. 8A. All the *C. tremulae* sequences as well as the other coleopteran ones share a glycoside hydrolase family 28 (GH28) conserved domain (IPR000743). The InterProScan search did not hit the PROSITE consensus pattern of GH28 proteins (PS00502) which is [GSDENKRH]-x(2)-[VMFC]-x(2)-[GS]-H-G-[LIVMAG]-x(1,2)-[LIVM]-G-S. This is because the active site consensus pattern of the coleopteran endopolygalacturonases is slightly different, we deduced it to be [SDENKRQT]-x(3)-C-x-G-[GS]-H-G-[LF]-x(3)-[IVTS]-G-x-S.

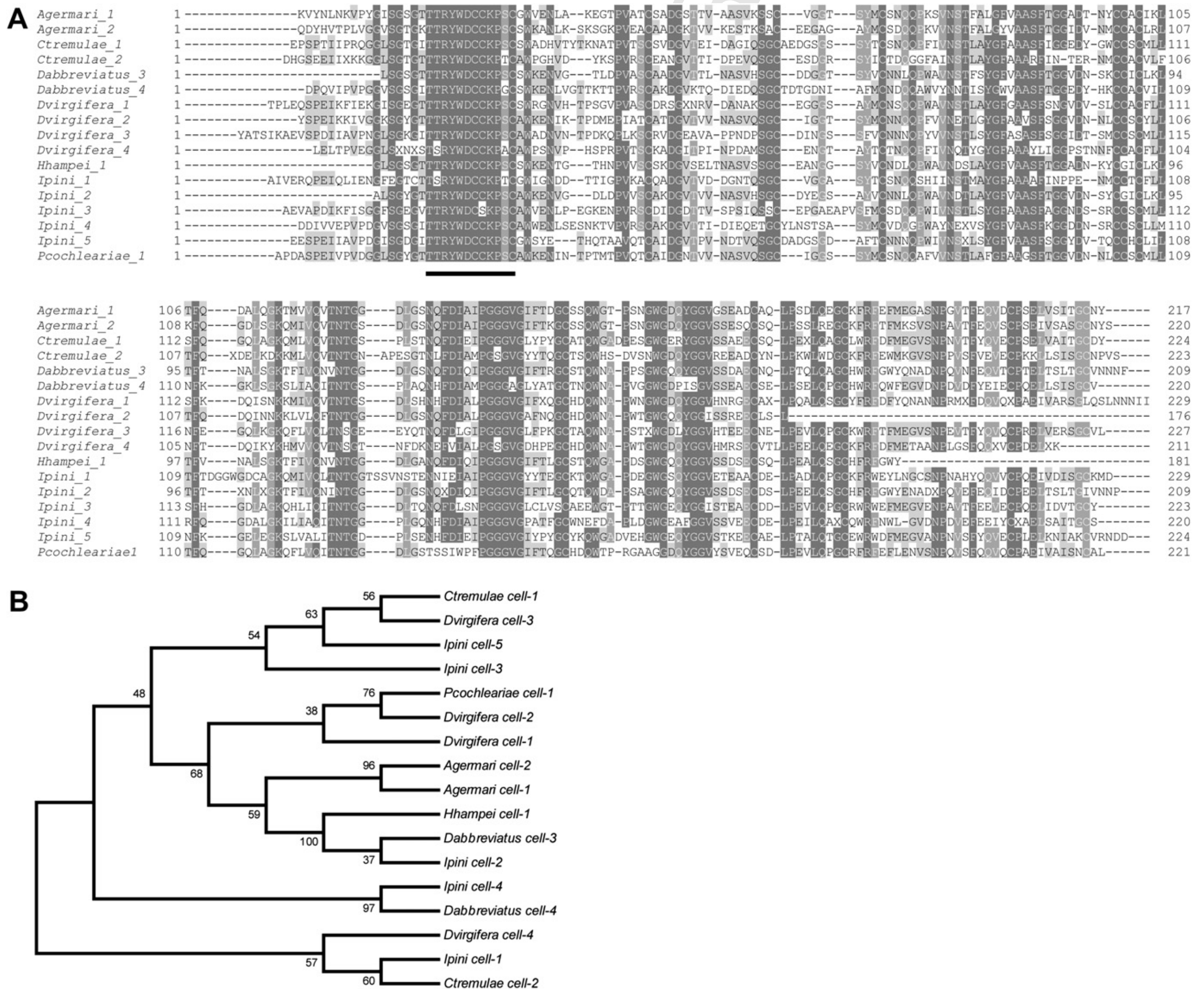


Fig. 7. Amino acid alignment and Phylogenetic analysis of the coleopteran cellulose gene family. (A) Predicted protein sequences deduced from coleopteran cellulase transcripts without their predicted signal peptide were aligned using MAFFT multiple alignment program and identical residues are boxed with dark shading, and conserved residues boxed with light shading. The predicted active site consensus pattern is underlined. (B) Phylogenetic analysis was conducted in MEGA4.0 using the UPGMA method and the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). Details on how these sequences were obtained are given as Supplementary materials (Table S3).

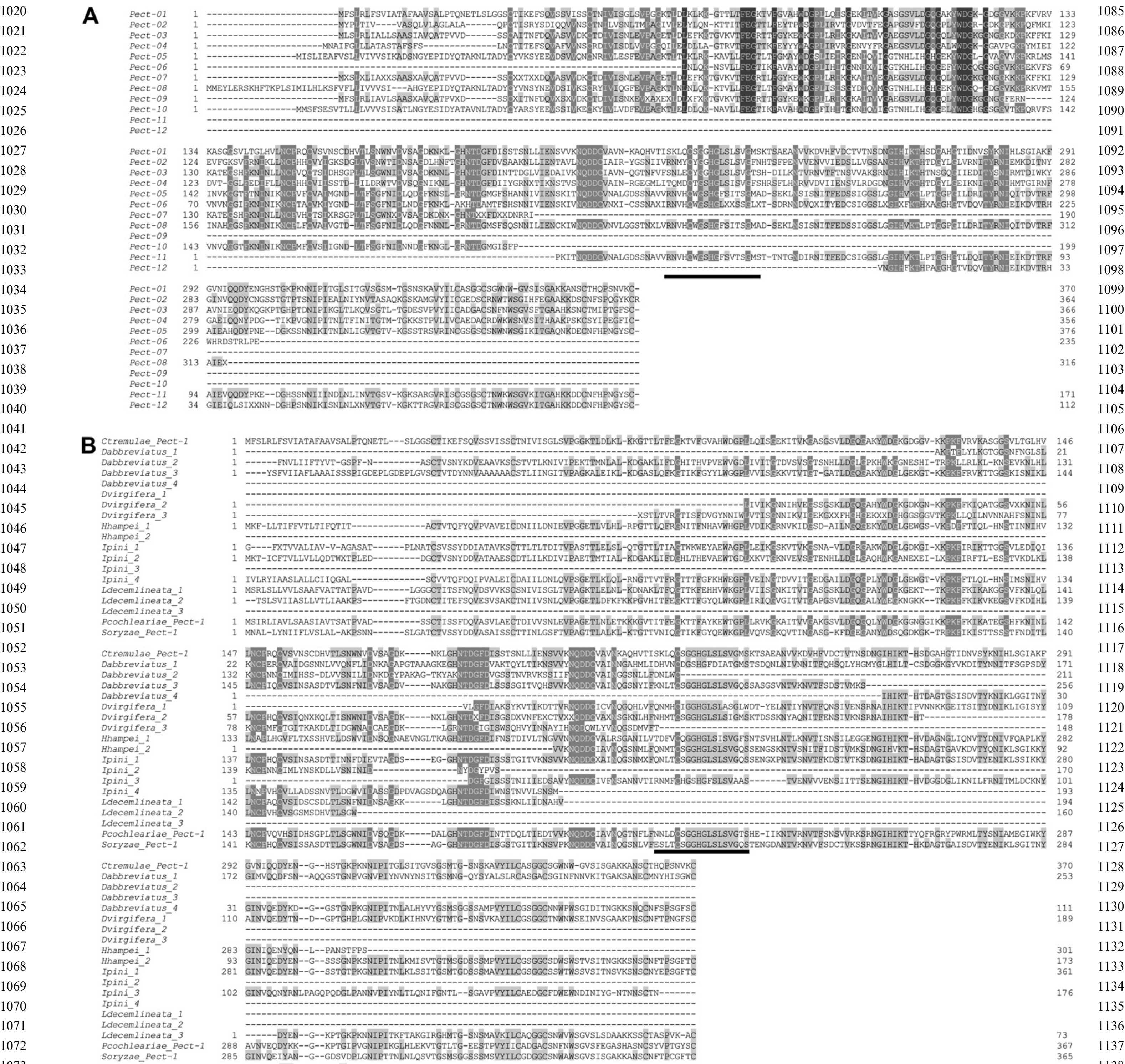


Fig. 8. Amino acid alignment of putative *C. tremulae* and other coleopteran endopolygalacturonases. Sequences were aligned using the MAFFT multiple alignment program and identical residues are boxed with dark shading, and conserved residues boxed with light shading. The predicted active site consensus pattern is underlined. Details on how these sequences were obtained are given as Supplementary materials (Table S3).

4. Discussion

Pyrosequencing looks set to dramatically increase the frequency and depth of EST collections for specific insect tissues and non-Drosophilid insects as a whole. Such datasets not only provide extensive nucleotide sequence data for insects lacking genomic data but also provide a plethora of genetic markers, such as microsatellites (Fig. 5), that can be used to derive high density linkage maps for mapping traits of biological relevance. In this fashion 454 sequencing can rapidly promote both sequence data and genetic markers in species for which no previous sequence data is available. For example, 454-mediated sequencing of cDNA from immune

related *M. sexta* tissues has already dramatically increased the number of ESTs available (Zou et al., 2008) and sequencing of other *Manduca* tissues, such as the midgut, will give us the first overview of a complete insect transcriptome from this physiological model which currently lacks any genomic framework. To our knowledge, the data described here is the first publication of a 454 derived insect midgut EST collection, and this dataset will also dramatically increase the number of midgut ESTs from collections previously derived via conventional capillary sequencing, such as those from the corn rootworm *D. virgifera* (Siegfried et al., 2005), the grass grub beetle *Costelytra zealandica* (Marshall et al., 2008) or the pine engraver *I. pini* (Eigenheer et al., 2003). Thus pyrosequencing of non-model insects is a dramatic and rapid way of providing genomic and transcriptomic resources in an organism where none are previously available. For example, with this publication, for *C. tremulae* the number of nucleotide sequences (reported contigs) has risen instantaneously from 8 (Genbank, 1st of February 2009) to over 23,000 (Table 1). When systematically applied to different tissues of insect models pyrosequencing will similarly dramatically increase resources available for all non-Drosophilids.

With specific reference to the midgut ESTs for *C. tremulae* described here, many of these predict proteins with functions expected for the midgut with likely roles in digestion, xenobiotic metabolism or immunity, whereas many either show unexpected functions or no similarity to predicted proteins in current databases. Expected roles in the midgut physiology of a leaf eating beetle involve destruction of the plant cell wall, thus the dominance of cellulases, pectinases and glucanases. However, what is perhaps more interesting is the absence of these predicted proteins from the *T. castaneum* genome. This predicted difference in the midgut enzyme composition of beetles feeding on stored products (*T. castaneum*) and fresh leaves (*C. tremulae*), begins to illustrate the potential power of comparing such datasets between insects with different lifestyles. We also note that enzymes degrading plant cell walls are of considerable current interest in biotechnology for the degradation of plant material for use in biofuels. 454-mediated sequencing of insect midguts, or salivary glands, may therefore be a powerful new way of finding enzymes important for plant biotechnology.

One of the central aims of this EST project was to identify contigs encoding candidate Bt receptors to use as probes in a dissection of the molecular basis of Cry3A resistance in *C. tremulae*. Contigs encoding three clear candidate Bt receptors were identified. First, a cadherin-like protein showing strong similarity to cadherin-like proteins from three other beetles, *T. castaneum*, *Tenebrio molitor* and *D. virgifera*. Second, eleven contigs encoding proteins similar to aminopeptidase N. Third, a contig encoding an alkaline phosphatase. However no contigs encoding ADAM10-like metalloproteases, recently reported as involved in the processing of Cry3A in Colorado potato beetle (Ochoa-Campuzano et al., 2007), were found.

Another expected function for proteins in the insect midgut is remodeling of peritrophic membrane (Table 3). The peritrophic membrane (PM) separates the food bolus in the insect midgut, in this case poplar leaves, from the midgut epithelium and associated brush border membrane. This membrane is composed of chitin and proteins and forms a permeable barrier that protects the midgut epithelium from erosion by the food bolus whilst allowing for the passage of enzymes and solutes involved in extracellular digestion in the gut (Caldeira et al., 2007). The PM requires continuous repair and replacement and it is therefore interesting to note dominance of enzymes involved in the binding (peritrophins, seven contigs), breakdown (chitinases, 36 contigs) and remodeling (chitin synthase, seven contigs, and chitin deacetylase, 13 contigs) of chitin (Table 3), suggesting that maintenance and remodeling of chitin in the PM occupies a major part of the *C. tremulae* transcriptional midgut machinery. It is interesting to note that a member of the group V chitin deacetylase was the most abundant transcript recovered from the *C. tremulae* non-normalized midgut cDNAs. This group of putative enzymes has attracted significant recent attention in several lepidopteran species (Campbell et al., 2008; Guo et al., 2005; Pauchet et al., 2008; Toprak et al., 2008), however their role in midgut physiology remains unclear.

As well as forming the major organ for processing of insect food, the midgut also forms the first line of defense against ingested xenobiotics and micro-organisms, the midgut therefore also plays a critical role in xenobiotic metabolism and insect immunity. With respect to xenobiotic metabolism, cytochrome P450s (25 hits) and glutathione-S-transferases (34 hits) are highly represented in the InterPro families detected (Table 2) and as full sequences for these gene families appear it will be interesting to compare the full complement of P450 and GST genes present in this leaf feeding beetle with that found in the grain feeding *T. castaneum*. For example, given that *C. tremulae* specializes in a single species of plant we might expect the range of enzymes required to metabolize secondary plant compounds to be reduced and somewhat specialized in their role.

In conclusion, here we have shown that 454-mediated pyrosequencing is a rapid and dramatic way of increasing the size and depth of insect midgut EST collections. Many of the predicted enzymes are currently only reported in the leaf beetle *C. tremulae*, sequenced here. Indeed some of the enzymes recovered are the first insect representatives. For example, the cellulose cellobiosidases (Table 3) appear to only have representatives from yeast, fungi and bacteria in current sequence databases. The approach taken here, of normalizing cDNAs from a specific insect tissue, attempts to define all of the unique transcripts associated with a tissue, often termed an 'unigene set'. We note, however, that cDNA shearing associated with preparation of the sample for 454 sequencing may introduce an element of bias resulting in uneven sequence coverage of representative cDNAs. However, the depth of EST coverage generated by 454 sequencing far outweighs that generated in conventional capillary led projects. 454 sequencing of other insect tissues will therefore not only increase the depth of existing capillary generated EST collections but will also identify new insect representatives for different enzymes and receptors. Further comparison of such datasets in insects with different lifestyles and food sources will begin to tell us how different insects exploit their different niches.

Acknowledgments

The authors thank Marcel Amichot for fruitful discussion, Claudine Courtin for insect rearing and Philippe Lorme for greenhouse plant production. Supported by a BBSRC grant to R. ff-C.

There is no conflict of interest.

Appendix. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ibmb.2009.04.001.

References

Augustin, S., Courtin, C., Rejasse, A., Lorme, P., Genissel, A., Bourguet, D., 2004. Genetics of resistance to transgenic *Bacillus thuringiensis* poplars in *Chrysomela tremulae* (Coleoptera: *Chrysomelidae*). J. Econ. Entomol. 97, 1058–1064.

Bowen, D., Rocheleau, T.A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R., ffrench-Constant, R.H., 1998. Insecticidal toxins from the bacterium *Photorhabdus luminescens*. Science 280, 2129–2132.

Buckingham, S.D., Biggin, P.C., Sattelle, B.M., Brown, L.A., Sattelle, D.B., 2005. Insect GABA receptors: splicing, editing, and targeting by antiparasitics and insecticides. Mol. Pharmacol. 68, 942–951.

Caldeira, W., Dias, A.B., Terra, W.R., Ribeiro, A.F., 2007. Digestive enzyme compartmentalization and recycling and sites of absorption and secretion along the midgut of *Dermestes maculatus* (Coleoptera) larvae. Arch. Insect Biochem. Physiol. 64, 1–18.

Campbell, P.M., Cao, A.T., Hines, E.R., East, P.D., Gordon, K.H., 2008. Proteomic analysis of the peritrophic matrix from the gut of the caterpillar, *Helicoverpa armigera*. Insect Biochem. Mol. Biol. 38, 950–958.
Conesa, A., Götz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008, 619832.
Dixit, R., Arakane, Y., Specht, C.A., Richard, C., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2008. Domain organization and phylogenetic analysis of proteins from the chitin deacetylase gene family of *Tribolium castaneum* and three other species of insects. Insect Biochem. Mol. Biol. 38, 440–451.
Dong, K., 2007. Insect sodium channels and insecticide resistance. Invert. Neurosci. 7, 17–30.
Eigenheer, A.L., Keeling, C.I., Young, S., Tittiger, C., 2003. Comparison of gene representation in midguts from two phytophagous insects, *Bombyx mori* and *Ips pini*, using expressed sequence tags. Gene 316, 127–136.
Federici, B.A., Bauer, L.S., 1998. Cyt1Aa protein of *Bacillus thuringiensis* is toxic to the cottonwood leaf beetle, *Chrysomela scripta*, and suppresses high levels of resistance to Cry3Aa. Appl. Environ. Microbiol. 64, 4368–4371.
ffrench-Constant, R.H., Daborn, P.J., Le Goff, G., 2004. The genetics and genomics of insecticide resistance. Trends Genet. 20, 163–170.
ffrench-Constant, R.H., Dowling, A., Waterfield, N.R., 2007. Insecticidal toxins from *Photorhabdus* bacteria and their potential use in agriculture. Toxicon 49, 436–451.
Genissel, A., Augustin, S., Courtin, C., Pilate, G., Lorme, P., Bourguet, D., 2003a. Initial frequency of alleles conferring resistance to *Bacillus thuringiensis* poplar in a field population of *Chrysomela tremulae*. Proc. R. Soc. Lond. B Biol. Sci. 270, 791–797.
Genissel, A., Leple, J.C., Millet, N., Augustin, S., Jouanin, L., Pilate, G., 2003b. High tolerance against *Chrysomela tremulae* of transgenic poplar plants expressing a synthetic cry3Aa gene from *Bacillus thuringiensis* ssp *tenebrionis*. Mol. Breed. 11, 103–110.
Gray, M.E., Sappington, T.W., Miller, N.J., Moeser, J., Bohn, M.O., 2009. Adaptation and invasiveness of western corn rootworm: intensifying research on a worsening pest. Annu. Rev. Entomol. 54, 303–321.
Guo, W., Li, G., Pang, Y., Wang, P., 2005. A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, *Trichoplusia ni*. Insect Biochem. Mol. Biol. 35, 1224–1234.
Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O.S., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., Gomez-Zurita, J., Ribera, I., Barraclough, T.G., Bocakova, M., Bocak, L., Vogler, A.P., 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science 318, 1913–1916.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
Marshall, S.D., Gatehouse, L.N., Becher, S.A., Christeller, J.T., Gatehouse, H.S., Hurst, M.R., Boucias, D.G., Jackson, T.A., 2008. Serine proteases identified from a *Costelytra zealandica* (White) (Coleoptera: Scarabaeidae) midgut EST library and their expression through insect development. Insect Mol. Biol. 17, 247–259.
Millar, N.S., Denholm, I., 2007. Nicotinic acetylcholine receptors: targets for commercially important insecticides. Invert. Neurosci. 7, 53–66.
Myhre, S., Tveit, H., Mollestad, T., Laegreid, A., 2006. Additional gene ontology structure for improved biological reasoning. Bioinformatics 22, 2020–2027.
Ochoa-Campuzano, C., Real, M.D., Martinez-Ramirez, A.C., Bravo, A., Rausell, C., 2007. An ADAM metalloprotease is a Cry3Aa *Bacillus thuringiensis* toxin receptor. Biochem. Biophys. Res. Commun. 362, 437–442.
Pauchet, Y., Muck, A., Svatos, A., Heckel, D.G., Preiss, S., 2008. Mapping the larval midgut lumen proteome of *Helicoverpa armigera*, a generalist herbivorous insect. J. Proteome Res. 7, 1629–1639.
Pigott, C.R., Ellar, D.J., 2007. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. Microbiol. Mol. Biol. Rev. 71, 255–281.
Raymond-Delpech, V., Matsuda, K., Sattelle, B.M., Rauh, J.J., Sattelle, D.B., 2005. Ion channels: molecular targets of neuroactive insecticides. Invert. Neurosci. 5, 119–133.
Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Beeman, R.W., Brown, S.J., Bucher, G., Friedrich, M., Grimmelikhuijzen, C.J., Klingler, M., Lorenzen, M., Richards, S., Roth, S., Schroder, R., Tautz, D., Zdobnov, E.M., Muzny, D., Gibbs, R.A., Weinstock, G.M., Attaway, T., Bell, S., Buhay, C.J., Chandrabose, M.N., Chavez, D., Clerk-Blankenburg, K.P., Cree, A., Dao, M., Davis, C., Chacko, J., Dinh, H., Dugan-Rocha, S., Fowler, G., Garner, T.T., Garnes, J., Gnirke, A., Hawes, A., Hernandez, J., Hines, S., Holder, M., Hume, J., Jhangiani, S.N., Joshi, V., Khan, Z.M., Jackson, L., Kovar, C., Kowis, A., Lee, S., Lewis, L.R., Margolis, J., Morgan, M., Nazareth, L.V., Nguyen, N., Okwuonu, G., Parker, D., Richards, S., Ruiz, S.J., Santibanez, J., Savard, J., Scherer, S.E., Schneider, B., Sodergren, E., Tautz, D., Vattahil, S., Villasana, D., White, C.S., Wright, R., Park, Y., Beeman, R.W., Lord, J., Oppert, B., Lorenzen, M., Brown, S., Wang, L., Savard, J., Tautz, D., Richards, S., Weinstock, G., Gibbs, R.A., Liu, Y., Worley, K., Weinstock, G., Elsik, C.G., Reese, J.T., Elhaik, E., Landan, G., Graur, D., Arensburger, P., Atkinson, P., Beeman, R.W., Beidler, J., Brown, S.J., Demuth, J.P., Drury, D.W., Du, Y.Z., Fujiwara, H., Lorenzen, M., Maselli, V., Osanai, M., Park, Y., Robertson, H.M., Tu, Z., Wang, J.J., Wang, S., Richards, S., Song, H., Zhang, L., Sodergren, E., Werner, D., Stanke, M., Morgenstern, B., Solovyev, V., Kosarev, P., Brown, G., Chen, H.C., Ermolaeva, O., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Maglott, D., Pruitt, K., Sapojnikov, V., Souvorov, A., Mackey, A.J., Waterhouse, R.M., Wyder, S., Zdobnov, E.M., Zdobnov, E.M., Wyder, S., Kriventseva, E.V., Kadowaki, T., Bork, P., Aranda, M., Bao, R., Beermann, A., Berns, N., Bolognesi, R., Bonneton, F., Bopp, D., Brown, S.J., Bucher, G., Butts, T., Chaumot, A., Denell, R.E., Ferrier, D.E., Friedrich, M., Gordon, C.M., Jindra, M., Klingler, M., Lan, Q., Lattorff, H.M., Laudet, V., von Levetsow, C., Liu, Z., Lutz, R., Lynch, J.A., da Fonseca, R.N., Posnien, N., Reuter, R., Roth, S., Savard, J., Schinko, J.B., Schmitt, C., Schoppmeier, M., Schroder, R., Shippy, T.D., Simonnet, F., Marques-Souza, H., Tautz, D., Tomoyasu, Y., Trauner, J., Van der Zee, M., Vervoort, M., Wittkopp, N., Wimmer, E.A., Yang, X., Jones, A.K., Sattelle, D.B., Ebert, P.R., Nelson, D., Scott, J.G., Beeman, R.W., Muthukrishnan, S., Kramer, K.J., Arakane, Y., Beeman, R.W., Zhu, Q., Hogenkamp, D., Dixit, R., Oppert, B., Jiang, H., Zou, Z., Marshall, J., Elpidina, E., Vinokurov, K., Oppert, C., Zou, Z., Evans, J., Lu, Z., Zhao, P., Sumathipala, N., Altincicek, B., Vilcinskas, A., Williams, M., Hultmark, D., Hetru, C., Jiang, H., Grimmelikhuijzen, C.J., Hauser, F., Cazzamali, G., Williamson, M., Park, Y., Li, B., Tanaka, Y., Predel, R., Neupert, S., Schachtner, J., Verleyen, P., Raible, F., Bork, P., Friedrich, M., Walden, K.K., Robertson, H.M., Angeli, S., Foret, S., Bucher, G., Schuetz, S., Maleszka, R., Wimmer, E.A., Beeman, R.W., Lorenzen, M., Tomoyasu, Y., Miller, S.C., Grossmann, D., Bucher, G., 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature 452, 949–955.
Sayed, A., Nekl, E.R., Siqueira, H.A., Wang, H.C., ffrench-Constant, R.H., Bagley, M., Siegfried, B.D., 2007. A novel cadherin-like gene from western corn rootworm, *Diabrotica virgifera virgifera* (Coleoptera: Chrysomelidae), larval midgut tissue. Insect Mol. Biol. 16, 591–600.
Siegfried, B.D., Waterfield, N., ffrench-Constant, R.H., 2005. Expressed sequence tags from *Diabrotica virgifera virgifera* midgut identify a coleopteran cadherin and a diversity of cathepsins. Insect Mol. Biol. 14, 137–143.
Soberon, M., Gill, S.S., Bravo, A., 2009. Signaling versus punching hole: how do *Bacillus thuringiensis* toxins kill insect midgut cells? Cell. Mol. Life Sci., doi:10.1007/s00018-008-8330-9.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Toprak, U., Baldwin, D., Erlandson, M., Gillott, C., Hou, X., Coutu, C., Hegedus, D.D., 2008. A chitin deacetylase and putative insect intestinal lipases are components of the *Mamestra configurata* (Lepidoptera: Noctuidae) peritrophic matrix. Insect Mol. Biol. 17, 573–585.
Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Vagner, L.L., Khaspekov, G.L., Kozhemyako, V.B., Matz, M.V., Meleshkevitch, E., Moroz, L.L., Lukyanov, S.A., Shagin, D.A., 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. 32, e37.
Zou, Z., Najar, F., Wang, Y., Roe, B., Jiang, H., 2008. Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. Insect Biochem. Mol. Biol. 38, 677–682.

UNCORRECTED PROOF