Overview of chitin metabolism enzymes in *Manduca sexta*: identification, domain organization, phylogenetic analysis and gene expression.

Guillaume Tetreau^a, Xiaolong Cao^b, Yun-Ru Chen^c, Subbaratnam Muthukrishnan^d, Jiang Haobo^b, Gary W. Blissard^c, Michael R. Kanost^d, Ping Wang^a.

^aDepartment of Entomology, Cornell University, New York State Agricultural Experiment Station, Geneva, NY 14456, USA

^bDepartment of Entomology and Plant Pathology, Oklahoma State University, Stillwater, OK 74078, USA

^cBoyce Thompson Institute, Cornell University, Ithaca, NY 14853-1801, USA

^dDepartment of Biochemistry & Molecular Biophysics, Kansas State University, 141 Chalmers Hall, Manhattan, KS 66506, USA

Abbreviations: CBD, chitin binding domain; CDA, chitin deacetylase; CE4, carbohydrate esterase family 4; CHS, chitin synthase; CHT, chitinase; PM, peritrophic membrane;

Abstract

Chitin is one of the most abundant biomaterials in nature. The biosynthesis and degradation of chitin in insects are complex and dynamically regulated to cope with insect growth and development. Chitin metabolism in insects is known to involve numerous enzymes, including chitin synthases (synthesis of chitin), chitin deacetylases (modification of chitin by deacetylation) and chitinases (degradation of chitin by hydrolysis). In this study, we conducted a genome-wide search and analysis of genes encoding these chitin metabolism enzymes in *M. sexta*. Our analysis confirmed that only two chitin synthases are present in *M. sexta* as in most other arthropods. Eleven chitin deacetylases (encoded by nine genes) were identified, with at least one representative in each of the five phylogenetic groups that have been described for chitin deacetylases to date. Eleven genes encoding for family 18 chitinases (GH18) were found in the *M. sexta* genome. Based on the presence of conserved sequence motifs in the catalytic sequences and phylogenetic relationships, two of the *M. sexta* chitinases did not cluster with any of the current eight phylogenetic groups of chitinases: two new groups were created (groups IX and X) and their characteristics are described. The result of the analysis of the Lepidoptera-specific chitinase-h (group h) is consistent with its proposed bacterial origin. By analyzing chitinases from fourteen species that belong to seven different phylogenetic groups, we reveal that the chitinase genes appear to have evolved sequentially in the arthropod lineage to achieve the current high level of diversity observed in *M. sexta*. Based on the sequence conservation of the catalytic domains and on their developmental stage- and tissue-specific expression, we propose putative functions for each group in each category of enzymes.

Keywords: chitin synthases; chitin deacetylases; chitinases; Lepidoptera; tobacco hornworm; RNA-seq; phylogenetic analysis

1. Introduction

Chitin, the second most abundant natural biopolymer on earth, is an important extracellular polysaccharide matrix found in the cell walls of fungi and diatoms, the shells of mollusks, egg shells of nematodes, and more abundantly in the exoskeleton of arthropods (Cohen, 2010; Rudall and Kenchington, 1973). In insects, chitin is a major structural component of the cuticle and of the peritrophic membrane (PM) that lines the midgut epithelium (Andersen, 1979; Rudall and Kenchington, 1973). In insect cuticles and the PM, chitin is associated with proteins and other components, such as catecholamines, lipids and metal ions, to form the unique chitinous structures (Kramer et al., 1988; Vincent and Wegst, 2004). The sclerotized cuticle serves as the exoskeleton that contributes to the evolutionary success of insects (Labandeira, 2005). The chitinous cuticle not only provides physical support, but also serves to prevent desiccation and to protect insects from physical and chemical injuries and pathogen infections. Metabolism of chitin in the cuticle and the PM is dynamically regulated to cope with growth, development and other physiological processes in insects. The rigidity of cuticle provides physical support and protection, but it restricts insect growth and development. Therefore, growth and development of insects are periodically interspersed with molting. Molting is a complex and multi-step cascade involving a series of enzymes and cofactors (Merzendorfer and Zimoch, 2003). Enzymes and proteins that are involved in the turnover or organization of chitin in the cuticle and PM modulate the physical properties of the chitin-containing structures and may affect the physiological status and survival of the insect (Arakane et al., 2009; Arakane et al., 2008). Therefore, chitin-related enzymes play fundamental roles in chitin metabolism, which is essential and critical for insect survival, growth and development.

Enzymes involved in chitin metabolism can be divided into three main categories, based on their functions to synthesize chitin (chitin synthases), to enzymatically alter chitin by

deacetylation (chitin deacetylases) and to degrade chitin by hydrolytic process (chitinases and *N*-acetylglucosaminidases).

The formation of chitin from monomers of UDP- *N*-acetylglucosamine is catalyzed by chitin synthases (CHS). Although the number of genes encoding chitin synthases (CHS) in different fungal species can range from 1 to >20, insect genomes characterized so far contain only two *CHS* genes (Merzendorfer, 2011). These two CHS have different tissue specificity of expression and it has been suggested that they have distinct but complementary functions in insect physiology: one is specifically expressed in the midgut epithelial cells for the synthesis of chitin for PM formation, while the other is involved in the synthesis of chitin for cuticle formation (Arakane et al., 2004). In the process of molting, the old cuticle is digested by specific enzymes such as chitinases and the new cuticle is secreted. In the formation of cuticles, chitin may be further modified by chitin deacetylases.

Chitin deacetylases (CDAs) are carbohydrate esterases that catalyze *N*-deacetylation of chitin to convert to deacetylated chitin. In insects and other arthropods, the CDAs bring about partial deacetylation of chitin (10-20% of GlcNAc residues in chitin are deacetylated) (Muzzarelli, 1977). The gene structure, domain organization and characteristics of CDAs have been described in insects (Arakane et al., 2009; Dixit et al., 2008). However, the functional roles of CDAs have been mostly studied and characterized in fungi (Zhao et al., 2010). The importance of some CDA genes in insects has been examined by RNAi and the results reported indicate that CDAs may play a critical role in molting, tracheal tubule development and possibly in insect immunity (Arakane et al., 2009; Dixit et al., 2008; Luschnig et al., 2006; Quan et al., 2013; Wang et al., 2006; Zhao et al., 2010). More recent studies in *Drosophila* have shown that CDAs are critically involved in tracheal morphogenesis (Dong et al., 2014; Luschnig et al., 2006). However, the biochemical characteristics of insect CDAs and the biochemical and

molecular mechanisms of the CDAs involved in structural formation, development and other physiological functions in insects are poorly understood at present.

Chitinases are a large family of enzymes that hydrolyze chitin into monomers of *N*-acetylglucosamine. They have been identified in many plants and animals, from protozoans to mammals including humans, and they are utilized mostly for defense and/or digestive functions (Adrangi and Faramarzi, 2013). Insect chitinases have a wide range of functions and differ substantially in their enzymatic properties, tissue and stage-specific gene expression, domain organization and size (Arakane and Muthukrishnan, 2010; Merzendorfer and Zimoch, 2003). The expression of some insect chitinase genes is tightly associated with molting, facilitating apolysis by degrading the chitin in the old cuticles, while some other chitinases are secreted into the midgut lumen presumably to regulate the PM structure and participate in the digestion of dietary chitin (Arakane and Muthukrishnan, 2010; Merzendorfer and Zimoch, 2003; Samuels and Reynolds, 1993; Zhu et al., 2008c). Even though chitinases have been identified in a wide range of organisms, little is known about the evolutionary history that has led to the high diversity of this protein family and their functions in insects (Funkhouser and Aronson, 2007). Understanding the diversity and evolution of this protein family should provide important new insights into insect physiology and metabolism. Moreover, studying chitinases can also have practical applications, notably by developing insect control strategies that target chitin metabolism, as chitinases are increasingly studied for their capacity to potentiate bioinsecticides (Gopalakrishnan et al., 1995; Kramer and Muthukrishnan, 1997) or to act as biocontrol agents (Assenga et al., 2006; Fitches et al., 2004; Kramer and Muthukrishnan, 1997; Rao et al., 2004). A comprehensive identification and analysis of enzymes involved in chitin metabolism will enrich the repertoire of insect molecular targets for the development of novel insect control strategies.

The tobacco hornworm *Manduca sexta* (Lepidoptera: Sphingidae) has been widely used as a model organism for studying insect physiology, biochemistry and molecular biology. The genome of *M. sexta* has been recently sequenced and assembled (<http://agripestbase.org/manduca> and http://www.ncbi.nlm.nih.gov/assembly/GCA_000262585.1). RNA sequencing data from different tissues and developmental stages are also available from several sources (<http://www.ncbi.nlm.nih.gov/bioproject/81039>). The present study is focused on the identification of chitin metabolism genes in the *M. sexta* genome and their expression patterns. We describe gene identification, sequence homology and domain organization of the chitin synthases, chitin deacetylases and chitinases from *M. sexta*. Based on phylogenetic analysis of these genes from *M. sexta* and other insect species and the patterns of gene expression in different tissues and at different developmental stages, we propose putative physiological functions for these enzymes in tobacco hornworm.

2. Materials and Methods

2.1. Identification of genes encoding chitin metabolism enzymes from the *M. sexta* genome

To identify the enzymes related to chitin metabolism in tobacco hornworm, we conducted an exhaustive sequence search of the *M. sexta* genome (Manduca base, <http://agripestbase.org/manduca>) for homologs of known chitin related enzyme genes. Chitin synthases, chitin deacetylases, and chitinases previously characterized from several lepidopteran and coleopteran species were obtained from the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov>) and used as queries to find homologs in the *M. sexta* genome by BLAST search. Then, the domain structure of these enzymes was determined by searching against the Conserved Domains Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler-Bauer et al., 2011). The sequences of the *M. sexta* chitin metabolism enzymes identified in the initial search were used as queries for a second round of “BLAST” search to identify additional chitin-related enzymes. This step was repeated until no additional genes were detected. In parallel, a BLAST search using the consensus catalytic domains of chitin synthases (Chitin_synth_C, [cd04190]), chitin deacetylases (CE4-like1, [cd10974] and CE4-like2, [cd10975]) and chitinases (glycoside hydrolase family 18 domains GH18, [pfam00704]) (see: <http://www.ncbi.nlm.nih.gov/cdd>, for more information on the conserved domains) as queries was performed to detect additional genes not found by the whole gene-to-gene search approach.

2.2 Phylogenetic analysis

2.2.1 Chitin deacetylases

A set of 72 proteins containing carbohydrate esterase 4 (CE4) domains (CE4-like1, [cd10974] and CE4-like2, [cd10975]), including 3 from Arachnida, 11 from Coleoptera, 13 from Crustacea, 11 from Diptera, 9 from Hymenoptera and 25 from Lepidoptera were

obtained from the NCBI database, from the Silkworm database (<http://www.silkbdb.org/>) or from the Manduca base and subjected to phylogenetic analysis (Supplementary Table 1). Sequences were aligned using ClustalW and then a phylogenetic tree was generated using the neighbor joining method (Poisson substitution model; uniform substitution rate; gaps/missing data treatment: complete deletion) with MEGA 5.10 software (Tamura et al., 2011). Robustness of the branches was assessed by performing a bootstrap analysis of 2000 replications. The chitin deacetylases detected in *M. sexta* genome were named MsCDA with a number corresponding to the phylogenetic group they belong to.

2.2.2 Chitinases

A set of 103 proteins containing glycoside hydrolase 18 domains (GH18, [smart00636]), including 23 from Coleoptera, 22 from Crustacea, 35 from Diptera, 1 from Echinodermata, 4 from Hymenoptera, 12 from Lepidoptera and 6 from Nematoda obtained from the NCBI database or from the Manduca base (Supplementary Table 2), were subjected to phylogenetic analysis using the maximum likelihood method (Jones-Taylor-Thornton model; Nearest-Neighbor-Interchange ML heuristic method) with MEGA 5.10 software (Tamura et al., 2011). Robustness of the branches was assessed by performing a bootstrap analysis of 2000 replications. The chitinases from *M. sexta* were named and numbered based on their positions in the phylogenetic tree and the group to which they belong: MsCHT (chitinases, all groups except group V) or MsIDGF (chitinase-like proteins – Imaginal Disk Growth Factors, group V).

2.3 Analysis of catalytic domains

2.3.1 Chitin deacetylases

Sequences of the catalytic domains from each CDA were compared and aligned using the ClustalW software from PBIL Expasy tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html) (Combet et al., 2000). The five conserved CE-4 family motifs implicated in catalysis described by Blair *et al.* (2006) and

Dixit *et al.* (2008) were searched. The presence of additional domains including a chitin binding domain (ChtBD2) and LDLa receptor domains in some of the CDAs were also identified by domain search analyses.

2.3.2 Chitinases

Organization of different domains (catalytic domain, transmembrane spans, etc.) was identified for all *M. sexta* chitinases and for one representative from other species for each group of chitinases. Domain organization was determined using the SMART tool (<http://smart.embl-heidelberg.de/>) by using the protein sequences as queries.

Alignment of the sequences of the catalytic domains from the chitinases of phylogenetic group IX from *M. sexta*, *T. castaneum*, *D. melanogaster* and *H. saltator* was performed as described in section 2.3.1. Same alignment was performed with group X chitinases from *M. sexta*, *D. plexippus*, *T. castaneum*, *C. floridanus*, *A. echinator* and *H. saltator*. For all chitinases, the motif FDxxDxDxE, typical of family 18 chitinases, was searched together with sequence signatures (regions with high sequence conservation) (Choi et al., 1997; de la Vega et al., 1998; Merzendorfer and Zimoch, 2003).

2.4 Developmental stage- and tissue-specific gene expression

Expression of chitin related genes was analyzed by RNA-seq from cDNA libraries prepared from different tissues at different developmental stages. RNA-seq data were acquired from the Manduca Genome project (http://www.ncbi.nlm.nih.gov/assembly/GCA_000262585.1). Official Gene Set (OGS) 1.0 was downloaded from <ftp://ftp.bioinformatics.ksu.edu/pub/Manduca/>. The chitin metabolism enzyme gene models in OGS1.0 were replaced by the manually verified sequences. Reads from different libraries were mapped to the OGS1.0 with updated chitin metabolism enzyme gene sequences using Bowtie software version 0.12.8 (Langmead et al., 2009), with all reads trimmed to 50 bp. The FPKM (fragments per

214 kilobase per million fragments mapped) values of each gene in each library were further
215 calculated by RSEM software version 1.2.12 (Li and Dewey, 2011).
216 A gene was considered to be expressed in a specific tissue or at a developmental stage
217 when its FPKM value in the corresponding library was higher than 30. This threshold was
218 chosen based on the finding that more than 90% of all *Manduca* genes from all the cDNA
219 libraries have FPKM below this value. Genes were considered as highly expressed when
220 their FPKM was higher than 200 (more than 98% of genes with FPKM below this value).
221

3. Results & Discussion

The exhaustive BLAST search of the genome of *M. sexta* yielded a total of 22 genes encoding 24 proteins that showed sequence similarity to the well characterized chitin metabolism enzymes. These genes putatively code for 2 chitin synthases, 11 chitin deacetylases and 11 chitinases (Table 1).

3.1 Chitin synthases

3.1.1 CHS identification and characterization

Only two chitin synthase genes were found in the *M. sexta* genome, which correspond to the two genes that had been previously identified: *MsCHS1* (Genbank accession number: AY062175) and *MsCHS2* (AY821560) (Table 1) (Hogenkamp et al., 2005; Zhu et al., 2002). This result is consistent with the observation that no more than two *CHS* genes have been identified in any insect species investigated so far (Merzendorfer, 2011).

3.1.2 CHS gene expression patterns

The RNA-seq data showed that *MsCHS1* was expressed in the head and abdomen of the *M. sexta* larvae and adults, tissues that include epidermal cells involved in cuticle production, while *MsCHS2* was exclusively expressed in the midgut (Table 2). The expression patterns of *MsCHS1* and *MsCHS2* genes are in agreement with previous reports that localized *MsCHS1* transcripts in the tracheae and epidermis and *MsCHS2* in the midgut of *M. sexta* larvae (Hogenkamp et al., 2005; Zimoch et al., 2005). The differential expression patterns of the two chitin synthase genes have been suggested to represent the different complementary functions of the two CHS in insects (Arakane et al., 2004; Hogenkamp et al., 2005): *MsCHS2*, secreted by the midgut epithelial cells, appears to be responsible for the synthesis of the chitin in the PM, while *MsCHS1* likely synthesizes chitin found in the epidermal cuticle as well as the tracheal cuticle.

3.2 Chitin deacetylases

3.2.1 CDA identification and characterization

In total, nine *CDA* genes were identified in the genome of *M. sexta* using known insect *CDA* sequences as queries (Table 1). Differential splicing was identified in two of the genes. Two alternative transcripts, named *MsCDA2A* and *MsCDA2B*, were identified as arising from the *MsCDA2* gene, and similarly two transcripts, *MsCDA5A* and *MsCDA5B*, were identified as transcripts from *MsCDA5* gene (Table 1). All twelve arthropod genomes that we analyzed had genes coding for chitin deacetylases containing the typical catalytic CE4 domain ([cd10974] and [cd10975], <http://www.ncbi.nlm.nih.gov/cdd>).

3.2.2 Phylogenetic and structural analysis of CDAs

Based on sequence homology and domain organization, the CDAs were divided into five groups (Dixit et al., 2008) (Fig. 1). CDAs from groups I to IV all contain a peritrophin-A type (ChtBD2) chitin-binding domain (CBD) in addition to the catalytic domain, while CDAs belonging to group V contain only the catalytic domain. The addition of chitin-binding domains may increase the affinity of CDAs to their substrate and facilitate their catalytic activity, although it has been demonstrated that some CDAs without CBD can bind to chitin (Guo et al., 2005). CDAs from groups I and II also contain an additional LDLa domain (low-density lipoprotein receptor class-A, [SM00192]). Although LDL receptors (with multiple LDLa repeats) are known to play a key role in mammalian cholesterol homeostasis, the role of such domains in insect CDAs remains unclear. One possibility is that CDA resorption into the cells lining tracheae involves this domain (Tsarouhas et al., 2007).

Two different catalytic domains were found among the five groups of CDAs. Groups I to III contain a CE4-like1 domain ([cd10974]) while groups IV and V contain a CE4-like 2 domain ([cd10975]). Although all the chitin deacetylases show similarities in their amino acid sequences and domain organization, the divergence of CE4-like1 and CE4-like2

domains are significantly different in the catalytic motifs (Fig. 2). The first catalytic motif (motif TFDD) is perfectly conserved in MsCDA1 and in all CDAs containing the CE4-like2 domain (groups IV and V). It contains the two aspartates that are critical for enzymatic activity by coordinating zinc and by binding the acetate released from the substrate (Blair et al., 2006), although no experimental evidence has been reported to support CDA activity of any insect CDA1 and CDA2 to date. In contrast, this domain is not well conserved in the other CDAs: MsCDA4 lacks the second aspartate (TFDG) while MsCDA2A, MsCDA2B and MsCDA3 lack both (TFNG). Such modifications of this first motif have also been described in *Tribolium castaneum*, *Apis mellifera*, *Drosophila melanogaster* and *Anopheles gambiae* (Dixit et al., 2008). It was proposed that the first aspartate is critical for the enzymatic activity (Blair et al., 2005), and that substitution of this aspartate could lead to loss of activity. The sequence of the second motif (H[S/T]I[S/T]H) is also well conserved in MsCDA1 and in CDAs from groups IV and V. It contains the two histidines that are crucial for enzyme activity by binding to a zinc atom (Fig. 2) (Blair et al., 2005). As observed for the first motif, the sequence of the second motif is not fully conserved in the four other CDAs, lacking the first histidine (MsCDA2A and MsCDA2B) or both histidines (MsCDA3 and MsCDA4). The third motif (RXP[F/Y]L), also responsible for zinc and aspartate binding, is well conserved in all the CDAs. The fourth motif is not well conserved and lacks the critical tryptophan residue (last amino acid) in all the CDAs while the fifth domain is well conserved, containing the leucine and/or the histidine residue required for acetate-binding. Interestingly, it has to be noted that the catalytic domains of MsCDA1 (group I) are more similar to those of CDAs from group IV and V, than to catalytic domains of the other members of the group I (MsCDA2A and MsCDA2B). Altogether, these results reveal that MsCDA1 and the CDAs containing the CE4-like2 domain have a good conservation of the five motifs, while the deviations in the sequences of the catalytic domains of the other CDAs could affect their chitin

deacetylase activity. Further functional studies to identify the enzymatic activity of the insect CDAs are required to verify if they have deacetylase activity and to better understand the role they could play in the modification of chitin in the insect cuticle.

3.2.3 CDA gene expression patterns

MsCDA1 and *MsCDA2*, belonging to group I, were highly expressed at all developmental stages, mainly in eggs, and in the head, abdomen and fat body of the insect (Table 3). Orthologs of these enzymes in *C. fumiferana*, *D. melanogaster* and *T. castaneum* have been demonstrated to exhibit the same pattern of expression in trachea and epidermis (Arakane et al., 2009; Luschnig et al., 2006; Quan et al., 2013). Expression of group I and group II CDAs in fat body in Lepidoptera has not been reported before and could reflect their expression in the trachea associated with the fat body. RNAi experiments performed in *C. fumiferana*, *Nilaparvata lugens* and *T. castaneum* suggested that group I CDAs play an important role in molting at all developmental stages, required for maintaining the body shape and cuticle integrity during the development of the insect, from neonate to adult (Arakane et al., 2009; Quan et al., 2013; Xi et al., 2014). Our observation of the expression pattern of *MsCDA1* and *MsCDA2* is consistent with the known expression pattern of their orthologs *DmCDA1* (*serpentine*) and *DmCDA2* (*vermiform*) in *Drosophila*. *MsCDA1* and *MsCDA2* are known to play critical roles in trachea morphogenesis (Dong et al., 2014; Luschnig et al., 2006). Whether *MsCDA1* and *MsCDA2* play similar roles in tracheal development requires to be understood.

CDAs belonging to groups II to IV have been extensively studied in *T. castaneum* in which RNAi experiments in larvae showed that their down-regulation did not appear to affect the molting process or survival (Arakane et al., 2009). However, RNAi of a group IV CDA expressed in the integument of *N. lugens* was found to lead to molting defects (Xi et al., 2014). *MsCDA3* (group II) was only expressed in the eggs (Table 3) while group II CDA was expressed in thoracic muscle and at the pharate adult stage of *T. castaneum*

(Arakane et al., 2009). *MsCDA4* (group III) exhibited a pattern of gene expression similar to that of genes of group I (Table 3). Among the group IV enzymes, *MsCDA5A* was expressed mainly in the head at all developmental stages, while *MsCDA5B* expression was not detected in any tissue and at any developmental stage (Table 3). This lack of expression of *MsCDA5B* may have been due to the pipeline used for the RNAseq reads which do not discriminate between the reads corresponding to the two alternatively spliced transcripts. The expression patterns of the MsCDAs of groups II to IV all differed from the ones observed in *T. castaneum* (Arakane et al., 2009). Even though they seem to lack the critical conserved sequences of their catalytic domains (Fig. 2), suggesting that these CDAs may not be active deacetylases, they may have other important functions in insects (such as immunity or digestion) because all insects (except hemimetabolous insects) studied so far have representatives in each of the five groups of CDAs. Their physiological functions need to be systematically investigated in multiple insect species under a variety of conditions to better understand their roles under laboratory versus field conditions.

Group V CDAs (*MsCDA6* to 9) were all found to be highly and exclusively expressed in the midgut of larvae (Table 3). The same pattern of expression has been described previously for group V CDAs from *T. castaneum*, *H. armigera* and *T. ni* (Arakane et al., 2009; Campbell et al., 2008; Guo et al., 2005; Pauchet et al., 2008). In *T. castaneum*, it has also been shown that their expression increased during the feeding stages and that there was an asymmetric distribution of the transcripts: *CDA6* was found expressed in anterior and posterior midgut, *CDA7* in posterior midgut exclusively and *CDA8* and 9 throughout the midgut of *T. castaneum* larvae (Arakane et al., 2009). In *M. sexta* also, the CDAs belonging to group V are expressed only in the midgut and during feeding stages only. RNAi experiments in *T. castaneum* revealed that the substantial down-regulation of transcripts of all group V CDAs genes simultaneously did not result in any noticeable

modification of insect morphology or development (Arakane et al., 2009). Therefore, the biochemical and physiological functions of the CDAs of this group remain unknown. However, in these CDAs, the highly conserved catalytically critical sequence motifs are the least divergent from the consensus motif for CDAs belonging to CE4-family esterases, suggesting that they are likely to be active enzymes. Whether these CDAs actually possess chitin deacetylase activity remains to be understood. A group V CDA from *T. ni* and another group V CDA from *H. armigera* were successfully expressed in insect cells using the baculovirus expression system. Both recombinant CDAs showed chitin binding activity, but they lacked detectable chitin deacetylase activity (Guo et al., 2005; Jakubowska et al., 2010). It has been reported that a group V CDA from *M. configurata*, expressed in *E. coli*, had chitin deacetylase activity (Toprak et al., 2008). However, the reported CDA activity was associated with a protein outside the range of the expected molecular weight of the recombinant CDA.

Group V CDAs are secreted in the midgut lumen but are not expressed in cuticle forming tissues (Guo et al., 2005), suggesting that they participate in functions other than cuticle formation and molting, and may play a role in the defense of insects against pathogens as suggested by Jakubowska *et al.* (2010). The number of group V CDAs can greatly vary between insect species, from one (Diptera) to four genes (*M. sexta*, *T. castaneum*), and only one of them has been reported in *T. ni* (Guo et al., 2005). However, the *T. ni* genome has not been fully sequenced yet. It appears that hemimetabolous insects including Hemiptera and Anaplura do not have representatives of group V CDAs. It is notable that these species do not have a pupal stage, in which these CDAs are expressed in other species (Xi et al., 2014). The exact role of these CDAs and their potential functional redundancy need to be investigated.

3.3 Chitinases

3.3.1 CHT identification

A total of eleven genes encoding family 18 chitinases (GH18) were identified in the genome of *M. sexta* (Table 1). Among the eleven CHT genes, one chitinase in Group I has been described previously (Genbank accession nos. L49234 and U02270 for the gene and corresponding cDNA, respectively) (Choi et al., 1997; Kramer et al., 1993). In addition, a cDNA containing an “hemocyte aggregation inhibitor protein” domain with several conserved sequence motifs of chitinases has also been previously reported (Genbank accession no. ACW82749) (Kanost et al., 1994). The protein encoded by this cDNA is placed in group IV along with *Drosophila* imaginal disk growth factors (IDGF), but presumably lacks chitinase activity as it misses the critical glutamate residue in the consensus sequence motif II (Arakane and Muthukrishnan, 2010; Zhu et al., 2008b). Another *M. sexta* cDNA that has been previously reported encodes a “chitinase-h” (Genbank accession no. ABB88891). This protein is an ortholog of the *Bombyx mori* protein, BmChi-h that has high sequence similarity to bacterial and baculovirus chitinases with an “N-terminal early set domain” involved in multimeric subunit interactions (Daimon et al., 2003).

In addition to these three previously described chitinases, our genome-wide search identified 8 additional genes coding for chitinases to bring the total number of *M. sexta* genes encoding chitinase-like proteins to 11. Nine of these chitinases exhibit the typical features characteristic of the eight groups previously described for chitinases (Arakane and Muthukrishnan, 2010), except MsCHT6 which is a member of group VI with an uncommon additional chitin-binding domain at its C-terminal region (Fig. 3). In addition to these eight groups, we identified two new genes encoding chitinases that diverged from the eight groups based on domain organization and sequences of conserved motifs, and they were clustered into two new groups with orthologous sequences from other species.

3.3.2 Phylogenetic and structural analysis of CHT

A BLAST search of insect genomes with the sequences of these two new members of the chitinase family as queries identified orthologs of these two proteins not only in a wide range of arthropods including, *Camponotus floridanus* (Cf), *D. melanogaster* (Dm), *Danaus plexippus* (Dpl), *Daphnia pulex* (Dp), *Harpegnathos saltator* (Hs), *T. castaneum* (Tc) but also in the nematode *C. elegans* (Ce) and the sea urchin, *Strongylocentrotus purpuratus* (Sp) (Fig. 4). Orthologs of these proteins have been previously detected in *Bombyx mori* (Pan et al., 2012), but here we report a comprehensive analysis of these new chitinase sequences across different species. These two new chitinases were named MsCHT1 and MsCHT3 and they belong to the two newly created groups IX and X, respectively. It is to be noted that two *Drosophila* chitinase genes, initially named *DmCHT1* and *DmCHT3* (de la Vega et al., 1998; Zhu et al., 2004), do not belong to these groups as they were later identified to be the catalytic domains of the larger *DmCHT10* gene of *Drosophila*, which belongs to group II (Zhu et al., 2008a).

The first group (group IX) includes several proteins previously identified as GH18 "Stabilin-1 interacting chitinase-like proteins" (GH18_SI-CLP; [cd08276]). These proteins are apparently involved in protein sorting during endocytosis and interact with the transmembrane receptor stabilin-1 and are secreted by lysosomes (Kzhyshkowska et al., 2006). Orthologs for this protein are found not only in insects but in almost all organisms of the protostomes and deuterostomes, including mammals (Funkhouser and Aronson, 2007).

Group IX chitinases have one catalytic domain but they lack the linker and any CBDs (Fig. 3). They show a different set of conserved residues in the four motifs located near the catalytic site or the substrate binding cleft (Arakane and Muthukrishnan, 2010; Varela et al., 2002) (Fig. 5, 6; Supplementary Fig. 1). The catalytic domain of GH18 chitinases is known to contain the catalytically critical consensus sequence consisting of

426 FDG(L/F)DLDWE(Y/F)P (Choi et al., 1997; de la Vega et al., 1998; Merzendorfer and
 427 Zimoch, 2003). Many of the chitinases in groups I to IV have at least one of this highly
 428 conserved active site motif, which contains the proton donor glutamate (residue E) (Fig.
 429 5) (Lu et al., 2002; Watanabe et al., 1993), and additional acidic amino acids, which are
 430 usually one or two aspartates which influence catalytic parameters (Merzendorfer and
 431 Zimoch, 2003). The consensus sequences for the group IX differ substantially from those
 432 in group I and group V chitinases. Most notably, they differ in consensus II being devoid
 433 of the acidic amino acids known to play a role in catalysis. While the consensus for group
 434 V IDGF-like proteins retains the first two acidic amino acids in the consensus
 435 FDGFDLDWEYP, the group IX proteins conserve only the first aspartic acid in this
 436 consensus and appear to lack the other acidic groups shown to influence catalytic activity
 437 (Lu et al., 2002; Zhang et al., 2002).

438 Chitinases from group X have a unique domain arrangement that differs from the
 439 chitinases of all other groups. They have an N-terminal signal peptide preceding the
 440 catalytic domain followed by two very closely spaced tandem chitin-binding domains and
 441 a very long (1369-1763 amino acids-long) C-terminal stretch ending with a third CBD (Fig.
 442 3). The long stretch in the middle has a high percentage of
 443 serine/threonine/proline/tyrosine with a potential to be glycosylated. In two of the three
 444 group X chitinases from Hymenoptera (*A. echinator* and *C. floridanus*), an additional
 445 complement component C1Q domain (SMART accession no. SM00110) was present at
 446 the C-terminal end of the protein but the function of this domain is not clear. This group of
 447 CHTs appears to be specific to arthropods, as no orthologs with multiple CBDs were
 448 detected in nematode, crustacean and echinoid genomes investigated in the present and
 449 previous studies (Colbourne et al., 2011; Sodergren et al., 2006;
 450 The *C.elegans*_Sequencing_Consortium, 1998). The catalytic domain of group X

chitinases appears to have suffered a truncation/alteration of the typical second conserved motif FDGLDLDWEYP to V(D/Q)G(L/V)EIN---- (Fig. 5 and 7) that has been demonstrated to contain residues crucial for chitinolytic activity of CHT (Huang et al., 2000; Lu et al., 2002) and appears missing the first conserved motif. However, catalytic domains from group X CHTs exhibit an overall good sequence conservation without any additional domain easily detectable (Fig. 7).

To be noted, group V chitinases have an inserted loop of 15-24 amino acid residues between β 4 strand and α 4 helix (Arakane and Muthukrishnan, 2010; Varela et al., 2002), which is not present in chitinases from group I to IV, some of which have been shown to be catalytically active (Zhu et al., 2008c). This might represent a functional change in binding specificity and/or catalytic activity of chitinase like proteins belonging to this group.

The chitinase MsCHT-h did not cluster with any of the other ten groups described but clustered with orthologs from other species in a group named “group h” in the present article (Fig. 4). Orthologs of this chitinase-h can be found in 13 different Lepidoptera species and in several bacteria, but not in other insect species from other orders, as revealed by a “BLAST” search (Supplementary Table 3). Therefore, this supports the hypothesis of a horizontal transfer of this chitinase from bacteria to insects, and more specifically to Lepidoptera (Daimon et al., 2003). MsCHT-h contains a catalytic domain and a polycystic kidney disease 1 PKD1 domain (Fig. 3). The catalytic domain in MsCHT-h contains the typical highly-conserved catalytic motif (Fig. 5), suggesting its chitinolytic activity in addition to its potential PKD1-related function.

3.3.3 Diversification of CHT families

Comparative analysis of amino acid sequences of chitinases from genomes of fourteen animal species belonging to seven different phylogenetic groups allows us to infer a hypothetical scenario of the evolution of the chitinase, from Echinodermata to

Lepidoptera, as shown in Fig. 8. It should be noted that chitinase gene evolution in different lineages involved different gene expansion, diversification and decline events (Funkhouser and Aronson, 2007). This analysis is focused on the lineage of Lepidoptera, so the chitinase genes from other lineages used in the analysis included only those orthologous to the 11 *M. sexta* chitinase genes. The chitinase genes in other lineages may have different evolutionary scenarios. For example, most nematodes have *CHT* genes belonging to groups III, VIII and IX (Fig. 4), but the *C. elegans* genome contains more than 40 different *CHT*-like genes coding for other groups of CHTs identified by searching the NCBI database (data not shown), which are not present in *M. sexta*, and therefore they are not included in this analysis.

One group of chitinases (group IX) was shared by all fourteen species analyzed and two other groups (groups III and VIII) were found in all species studied except Echinodermata. Groups VIII and IX chitinases have a very simple domain organization with only one catalytic domain and no other detectable domains (Fig. 3). Chitinases from group III appeared in a common ancestor to Nematoda and other arthropods but the former has a simplified domain organization when compared to the arthropod ones, with only one catalytic domain and one (*B. malayi*) or two (*A. ceylanicum*, *C. elegans* and *H. contortus*) chitin-binding domains (Fig. 9). The domain organization of the group III chitinases in *M. sexta* (two catalytic domains and one chitin-binding domain) is shared by all the arthropods and crustaceans included in this study (Fig. 9). A major evolutionary event allowed the appearance of more complex chitinases from the groups I, II, IV, VI and VII in crustaceans and arthropods. Chitinase-h had then been acquired only by Lepidoptera by a horizontal transfer from bacteria. This group has not been found in other insect orders so far. Altogether, these results suggest that the chitinase family has evolved sequentially to achieve the high level of functional diversity in arthropods, notably in Lepidoptera. These results are in accordance with the previous model of evolution of the chitinase

family GH18, which suggested that an important expansion of chitinase genes occurred in protostomes, such as *C. elegans* and *D. melanogaster*, due to the importance of chitinous structures for protection and architectural features (Funkhouser and Aronson, 2007). An extensive sequence analysis would be necessary, by taking into account additional representative species in each phylogenetic group, to refine this model of CHT evolution.

3.3.4 CHT gene expression patterns

Unlike CHSs and CDAs, CHTs exhibit unique patterns of expression specific to tissues and/or developmental stages (Table 4). Majority of the chitinases are expressed in abdomen consistent with their involvement in the turnover of body cuticle. The midgut tissue expresses only a limited number of chitinase genes in larval (presumably feeding) stages. These genes may be involved in structural regulation of PM-associated chitin and possibly tracheal cuticle digestion. The exceptions may be group V chitinases, which have been called imaginal disk growth factors that seem to be expressed at all developmental stages and in all tissues, which is consistent with previous findings in *B. mori* (Pan et al., 2010). The lepidopteran-specific chitinase-h, which was suggested to have been acquired by horizontal transmission from a microbial source (Daimon et al., 2003), is also expressed more widely than the chitinases belonging to the other groups. *MsCHT1* has been found to be expressed exclusively in testes and ovaries (Table 4). Its reproductive tissue-specific expression may explain why this enzyme has not been described previously. A recent work revealed the presence of a chitin-like component in *Aedes aegypti* ovaries, during oogenesis, without characterizing its exact function (Moreira et al., 2007). Additional experiments are required to determine if *MsCHT1* is an active enzyme and if it modulates the effect of the chitin in the reproductive organs. The RNA-seq data showed that *MsCHT3* was expressed in the head and abdomen of fourth instar larvae (Table 4), but the specific tissue distributions and physiological roles of *MsCHT3* have yet to be studied in detail.

4. Conclusions

Proteins involved in chitin metabolism are key enzymes that synthesize, modify or degrade chitin in the cuticle or the PM facilitating growth of the insect through multiple molts and morphogenesis. A genome-wide analysis of the proteins provides the first extensive overview of the three main categories of chitin metabolism enzymes in the *M. sexta* genome. We detected 22 genes encoding 24 chitin metabolism enzymes in the genome of *M. sexta*, including 2 chitin synthases, 11 chitin deacetylases and 11 chitinases. Two chitin synthases were detected in *M. sexta*, which is consistent with previous reports. Among the CDAs, groups I and V CDAs have gene expression profiles and domain organization congruent with previous work. Although some chitinase groups found in insects are shared with Echinodermata and Nematoda, we demonstrated that several groups and paralogs have appeared and evolved sequentially throughout the evolutionary tree, in the arthropod branch where chitin began to have a major role in ensuring their evolutionary success. Two new phylogenetic groups of chitinases (groups IX and X) have been created in addition to the eight groups described previously, with group X being an arthropod-specific group of chitinases. Our analysis supports the hypothesis that chitinase-h has been acquired by horizontal transfer from bacteria and that it is present only in Lepidoptera but not in other insect orders or arthropods. Additional experiments are required to determine which of these chitin-related enzymes are enzymatically active and to characterize their functions in chitin metabolism in insects.

Acknowledgement

The authors thanks Dr. Alexie Papanicolaou for his help in expression data analysis during early version of the manuscript. This project was supported in part by the Cornell University Agricultural Experiment Station federal formula funds received from the USDA Cooperative State Research, Education, and Extension Service to PW, and by the

555 National Science Foundation through grant IOS-1022227 to SM. This is contribution no.
556 15-221-J from the Kansas Agricultural Experiment Station. The computing for this project
557 was performed at the OSU High Performance Computing Center (OSUHPCC) at
558 Oklahoma State University (OSU), which was supported in part through instrumentation
559 funded by the National Science Foundation through grant OCI-1126330. Genome and
560 RNA-seq data were provided by the Manduca Genome Project which was funded by a
561 Defense Advanced Research Projects Agency grant to GB and a National Institutes of
562 Health (GM041247) grant to MK.

563

564 References

- 565 Adrangi, S., Faramarzi, M.A., 2013. From bacteria to human: a journey into the world of
566 chitinases. *Biotechnology advances* 31, 1786-1795.
- 567 Andersen, S.O., 1979. Biochemistry of insect cuticle. *Annual Review of Entomology* 24,
568 24-61.
- 569 Arakane, Y., Dixit, R., Begum, K., Park, Y., Specht, C.A., Merzendorfer, H., Kramer, K.J.,
570 Muthukrishnan, S., Beeman, R.W., 2009. Analysis of functions of the chitin
571 deacetylase gene family in *Tribolium castaneum*. *Insect Biochemistry and*
572 *Molecular Biology* 39, 355-365.
- 573 Arakane, Y., Hogenkamp, D.G., Zhu, Y.C., Kramer, K.J., Specht, C.A., Beeman, R.W.,
574 Kanost, M.R., Muthukrishnan, S., 2004. Characterization of two chitin synthase
575 genes of the red flour beetle, *Tribolium castaneum*, and alternate exon usage in
576 one of the genes during development. *Insect Biochemistry and Molecular Biology*
577 34, 291-304.
- 578 Arakane, Y., Muthukrishnan, S., 2010. Insect chitinase and chitinase-like proteins.
579 *Cellular and Molecular Life Sciences* 67, 201-216.
- 580 Arakane, Y., Specht, C.A., Kramer, K.J., Muthukrishnan, S., Beeman, R.W., 2008. Chitin
581 synthases are required for survival, fecundity and egg hatch in the red flour beetle,
582 *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology* 38, 959-962.
- 583 Assenga, S.P., You, M., Shy, C.H., Yamagishi, J., Sakaguchi, T., Zhou, J.L., Kibe, M.K.,
584 Xuan, X.N., Fujisaki, K., 2006. The use of a recombinant baculovirus expressing a
585 chitinase from the hard tick *Haemaphysalis longicornis* and its potential application
586 as a bioacaricide for tick control. *Parasitology Research* 98, 111-118.
- 587 Blair, D.E., Hekmat, O., Schuttelkopf, A.W., Shrestha, B., Tokuyasu, K., Withers, S.G.,
588 van Aalten, D.M.F., 2006. Structure and mechanism of chitin deacetylase from the
589 fungal pathogen *Colletotrichum lindemuthianum*. *Biochemistry* 45, 9416-9426.
- 590 Blair, D.E., Schuttelkopf, A.W., MacRae, J.I., van Aalten, D.M.F., 2005. Structure and
591 metal-dependent mechanism of peptidoglycan deacetylase, a streptococcal
592 virulence factor. *Proceedings of the National Academy of Sciences of the United*
593 *States of America* 102, 15429-15434.
- 594 Campbell, P.M., Cao, A.T., Hines, E.R., East, P.D., Gordon, K.H.J., 2008. Proteomic
595 analysis of the peritrophic matrix from the gut of the caterpillar, *Helicoverpa*
596 *armigera*. *Insect Biochemistry and Molecular Biology* 38, 950-958.
- 597 Choi, H.K., Choi, K.H., Kramer, K.J., Muthukrishnan, S., 1997. Isolation and
598 characterization of a genomic clone for the gene of an insect molting enzyme,
599 chitinase. *Insect Biochemistry and Molecular Biology* 27, 37-47.
- 600 Cohen, E., 2010. Chitin Biochemistry: Synthesis, Hydrolysis and Inhibition, in: Casas, J.,
601 Simpson, S.J. (Eds.), *Advances in Insect Physiology: Insect Integument and*
602 *Colour*, pp. 5-74.
- 603 Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H.,
604 Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Caceres, C.E.,
605 Carmel, L., Casola, C., Choi, J.-H., et al., 2011. The Ecoresponsive Genome of
606 *Daphnia pulex*. *Science* 331, 555-561.
- 607 Combet, C., Blanchet, C., Geourjon, C., Deleage, G., 2000. NPS@: Network Protein
608 Sequence Analysis. *Trends in Biochemical Sciences* 25, 147-150.
- 609 Daimon, T., Hamada, K., Mita, K., Okano, K., Suzuki, M.G., Kobayashi, M., Shimada, T.,
610 2003. A *Bombyx mori* gene, *BmChi-h*, encodes a protein homologous to bacterial
611 and baculovirus chitinases. *Insect Biochemistry and Molecular Biology* 33, 749-
612 759.

- de la Vega, H., Specht, C.A., Liu, Y., Robbins, P.W., 1998. Chitinases are a multi-gene family in *Aedes*, *Anopheles* and *Drosophila*. *Insect Molecular Biology* 7, 233-239.
- 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 Biochemistry and Molecular Biology* 38, 440-451.
- Dong, B., Miao, G., Hayashi, S., 2014. A fat body-derived apical extracellular matrix enzyme is transported to the tracheal lumen and is required for tube morphogenesis in *Drosophila*. *Development* 141, 4104-4109.
- Fitches, E., Wilkinson, H., Bell, H., Bown, D.P., Gatehouse, J.A., Edwards, J.P., 2004. Cloning, expression and functional characterisation of chitinase from larvae of tomato moth (*Lacanobia oleracea*): a demonstration of the insecticidal activity of insect chitinase. *Insect Biochemistry and Molecular Biology* 34, 1037-1050.
- Funkhouser, J.D., Aronson, N.N., Jr., 2007. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evolutionary Biology* 7.
- Gopalakrishnan, B., Muthukrishnan, S., Kramer, K.J., 1995. Baculovirus-mediated expression of a *Manduca sexta* chitinase gene - properties of the recombinant protein. *Insect Biochemistry and Molecular Biology* 25, 255-265.
- Guo, W., Li, G.X., Pang, Y., Wang, P., 2005. A novel chitin-binding protein identified from the peritrophic membrane of the cabbage looper, *Trichoplusia ni*. *Insect Biochemistry and Molecular Biology* 35, 1224-1234.
- Hogenkamp, D.G., Arakane, Y., Zimoch, L., Merzendorfer, H., Kramer, K.J., Beeman, R.W., Kanost, M.R., Specht, C.A., Muthukrishnan, S., 2005. Chitin synthase genes in *Manduca sexta*: characterization of a gut-specific transcript and differential tissue expression of alternately spliced mRNAs during development. *Insect Biochemistry and Molecular Biology* 35, 529-540.
- Huang, X., Zhang, H., Zen, K.C., Muthukrishnan, S., Kramer, K.J., 2000. Homology modeling of the insect chitinase catalytic domain-oligosaccharide complex and the role of a putative active site tryptophan in catalysis. *Insect Biochemistry and Molecular Biology* 30, 107-117.
- Jakubowska, A.K., Caccia, S., Gordon, K.H., Ferre, J., Herrero, S., 2010. Downregulation of a Chitin Deacetylase-Like Protein in Response to Baculovirus Infection and Its Application for Improving Baculovirus Infectivity. *Journal of Virology* 84, 2547-2555.
- Kanost, M.R., Zepp, M.K., Ladendorff, N.E., Andersson, L.A., 1994. Isolation and characterization of a hemocyte aggregation inhibitor from hemolymph of *Manduca sexta* larvae. *Archives of Insect Biochemistry and Physiology* 27, 123-136.
- Kramer, K.J., Corpuz, L., Choi, H.K., Muthukrishnan, S., 1993. Sequence of a cDNA and expression of the gene encoding epidermal and gut chitinases of *Manduca sexta*. *Insect Biochemistry and Molecular Biology* 23, 691-701.
- Kramer, K.J., Hopkins, T.L., Schaefer, J., 1988. Insect Cuticle Structure and Metabolism, in: Hedin, P.A., Menn, J.J., Hollingworth, R.M. (Eds.), *Biotechnology for Crop Protection*. ACS Publications, pp. 160-185.
- Kramer, K.J., Muthukrishnan, S., 1997. Insect chitinases: Molecular biology and potential use as biopesticides. *Insect Biochemistry and Molecular Biology* 27, 887-900.
- Kzhyshkowska, J., Mamidi, S., Gratchev, A., Kremmer, E., Schmuttermair, C., Krusell, L., Haus, G., Utikal, J., Schledzewski, K., Scholtze, J., Goerdts, S., 2006. Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-regulated in alternatively activated macrophages and secreted via lysosomal pathway. *Blood* 107, 3221-3228.

- Labandeira, C.C., 2005. Invasion of the continents: cyanobacterial crusts to tree-inhabiting arthropods. *Trends in Ecology & Evolution* 20, 253-262.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10.
- Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12.
- Lu, Y.M., Zen, K.C., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed mutagenesis and functional analysis of active site acidic amino acid residues D142, D144 and E146 in *Manduca sexta* (tobacco hornworm) chitinase. *Insect Biochemistry and Molecular Biology* 32, 1369-1382.
- Luschnig, S., Batz, T., Armbruster, K., Krasnow, M.A., 2006. *serpentine* and *vermiform* encode matrix proteins with chitin binding and deacetylation domains that limit tracheal tube length in *Drosophila*. *Current Biology* 16, 186-194.
- Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., et al., 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39, D225-D229.
- Merzendorfer, H., 2011. The cellular basis of chitin synthesis in fungi and insects: Common principles and differences. *European Journal of Cell Biology* 90, 759-769.
- Merzendorfer, H., Zimoch, L., 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology* 206, 4393-4412.
- Moreira, M.F., dos Santos, A.S., Marotta, H.R., Mansur, J.F., Ramos, I.B., Machado, E.A., Souza, G.H.M.F., Eberlin, M.N., Kaiser, C.R., Kramer, K.J., Muthukrishnan, S., Vasconcellos, A.M.H., 2007. A chitin-like component in *Aedes aegypti* eggshells, eggs and ovaries. *Insect Biochemistry and Molecular Biology* 37, 1249-1261.
- Muzzarelli, R.A.A., 1977. Chitin. Pergamon Press, London.
- Pan, Y., Chen, K., Xia, H., Yao, Q., Gao, L., Lue, P., Huojuan, He, Y., Wang, L., 2010. Molecular Cloning, Expression and Characterization of BmIDGF Gene from *Bombyx mori*. *Zeitschrift Fur Naturforschung Section C-a Journal of Biosciences* 65, 277-283.
- Pan, Y., Lu, P., Wang, Y., Yin, L., Ma, H., Ma, G., Chen, K., He, Y., 2012. *In silico* identification of novel chitinase-like proteins in the silkworm, *Bombyx mori*, genome. *Journal of Insect Science* 12.
- 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. *Journal of Proteome Research* 7, 1629-1639.
- Quan, G., Ladd, T., Duan, J., Wen, F., Doucet, D., Cusson, M., Krell, P.J., 2013. Characterization of a spruce budworm chitin deacetylase gene: Stage- and tissue-specific expression, and inhibition using RNA interference. *Insect Biochemistry and Molecular Biology*.
- Rao, R., Fiandra, L., Giordana, B., de Eguileor, M., Congiu, T., Burlini, N., Arciello, S., Corrado, G., Pennacchio, F., 2004. AcMNPV ChiA protein disrupts the peritrophic membrane and alters midgut physiology of *Bombyx mori* larvae. *Insect Biochemistry and Molecular Biology* 34, 1205-1213.
- Rudall, K.M., Kenchington, W., 1973. The chitin system. *Biological Review* 49, 597-636.

- Samuels, R.I., Reynolds, S.E., 1993. Molting fluid enzymes of the tobacco hornworm, *Manduca sexta* - timing of proteolytic and chitinolytic activity in relation to pre-ecdysial development. *Archives of Insect Biochemistry and Physiology* 24, 33-44.
- Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., Coffman, J.A., Dean, M., Elphick, M.R., Ettensohn, C.A., Foltz, K.R., et al., 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314, 941-952.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28, 2731-2739.
- The *C.elegans* Sequencing Consortium, 1998. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* 282, 2012-2018.
- 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 Molecular Biology* 17, 573-585.
- Tsarouhas, V., Senti, K.-A., Jayaram, S.A., Tiklova, K., Hemphala, J., Adler, J., Samakovlis, C., 2007. Sequential pulses of apical epithelial secretion and endocytosis drive airway maturation in *Drosophila*. *Developmental Cell* 13, 214-225.
- Varela, P.F., Llera, A.S., Mariuzza, R.A., Tormo, J., 2002. Crystal structure of imaginal disc growth factor-2 - A member of a new family of growth-promoting glycoproteins from *Drosophila melanogaster*. *J. Biol. Chem.* 277, 13229-13236.
- Vincent, J.F., Wegst, U.G., 2004. Design and mechanical properties of insect cuticle. *Arthropod structure & development* 33, 187-199.
- Wang, S.Q., Jayaram, S.A., Hemphala, J., Senti, K.A., Tsarouhas, V., Jin, H.N., Samakovlis, C., 2006. Septate-junction-dependent luminal deposition of chitin deacetylases restricts tube elongation in the *Drosophila* trachea. *Current Biology* 16, 180-185.
- Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M., Tanaka, H., 1993. Identification of glutamic acid-204 and aspartic acid-200 in chitinase-A1 of *Bacillus-circulans* WL-12 as essential residues for chitinase activity. *J. Biol. Chem.* 268, 18567-18572.
- Xi, Y., Pan, P.-L., Ye, Y.-X., Yu, B., Zhang, C.-X., 2014. Chitin deacetylase family genes in the brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae). *Insect Molecular Biology*.
- Zhang, H., Huang, X., Fukamizo, T., Muthukrishnan, S., Kramer, K.J., 2002. Site-directed mutagenesis and functional analysis of an active site tryptophan of insect chitinase. *Insect Biochemistry and Molecular Biology* 32, 1477-1488.
- Zhao, Y., Park, R.-D., Muzzarelli, R.A.A., 2010. Chitin Deacetylases: Properties and Applications. *Marine Drugs* 8, 24-46.
- Zhu, Q., Arakane, Y., Banerjee, D., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008a. Domain organization and phylogenetic analysis of the chitinase-like family of proteins in three species of insects. *Insect Biochemistry and Molecular Biology* 38, 452-466.
- Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008b. Characterization of recombinant chitinase-like proteins of *Drosophila melanogaster* and *Tribolium castaneum*. *Insect Biochemistry and Molecular Biology* 38, 467-477.

763 Zhu, Q., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S., 2008c. Functional
 764 specialization among insect chitinase family genes revealed by RNA interference.
 765 Proceedings of the National Academy of Sciences of the United States of America
 766 105, 6650-6655.
 767 Zhu, Q.S., Deng, Y.P., Vanka, P., Brown, S.J., Muthukrishnan, S., Kramer, K.J., 2004.
 768 Computational identification of novel chitinase-like proteins in the *Drosophila*
 769 *melanogaster* genome. Bioinformatics 20, 161-169.
 770 Zhu, Y.C., Specht, C.A., Dittmer, N.T., Muthukrishnan, S., Kanost, M.R., Kramer, K.J.,
 771 2002. Sequence of a cDNA and expression of the gene encoding a putative
 772 epidermal chitin synthase of *Manduca sexta*. Insect Biochemistry and Molecular
 773 Biology 32, 1497-1506.
 774 Zimoch, L., Hogenkamp, D.G., Kramer, K.J., Muthukrishnan, S., Merzendorfer, H., 2005.
 775 Regulation of chitin synthesis in the larval midgut of *Manduca sexta*. Insect
 776 Biochemistry and Molecular Biology 35, 515-527.

777

778

Tables

Table 1. List of the chitin deacetylases, chitinases and chitin synthases detected in the genome of *M. sexta*.

	Gene name / Splice variant	Manduca Base Accession no.	GenBank Accession no.	Protein length (aa)	M.W. ^a (kDa)	pI ^a	CBD	Citations
Chitin deacetylases	MsCDA1	Msex2.03153		539	61.5	5.0	1	
	MsCDA2A	Msex2.03152		536	60.8	5.3	1	
	MsCDA2B	Msex2.03152		542	61.5	5.3	1	
	MsCDA3	Msex2.02612		525	61.0	5.8	1	
	MsCDA4	Msex2.02671		494	56.0	5.5	1	
	MsCDA5A	Msex2.03714		888	100.1	6.6	1	
	MsCDA5B	Msex2.03714		888	100.3	6.8	1	
	MsCDA6	Msex2.08537		381	43.5	5.7	0	
	MsCDA7	Msex2.08536		375	42.6	5.4	0	
	MsCDA8	Msex2.08538		392	44.7	5.5	0	
	MsCDA9	Msex2.08539		372	43.2	5.6	0	
Chitinases	MsCHT1	Msex2.02611		394	45.2	8.9	0	
	MsCHT2	Msex2.03860		453	50.9	6.2	0	
	MsCHT3	Msex2.04654		2271	247.6	6.1	3	
	MsCHT5	Msex2.08301	U02270	554	62.2	5.3	1	(Kramer et al., 1993)
	MsCHT6	Msex2.07800		3208	359.7	5.4	1	
	MsCHT7	Msex2.03556		996	113.4	6.3	1	
	MsCHT8	Msex2.10627		1181	127.9	4.8	1	
	MsCHT10	Msex2.04970		2898	321.9	5.9	7	
	MsCHT11	Msex2.12654		490	55.4	6.4	0	
	MsCHT-h	Msex2.02783	ABB88891	555	61.3	5.5	0	
Chitinase-like protein	MsIDGF1	Msex2.03419	ACW82749	434	48.2	7.1	0	(Kanost et al., 1994)
Chitin synthases	MsCHS1	Msex2.14083	AY062175	1563	178.6	6.6	0	(Zhu et al., 2002)
	MsCHS2	Msex2.06544	AY821560	1524	174.4	6.0	0	(Hogenkamp et al., 2005)

Footnotes:

^aMolecular weight (M.W.) and isoelectric point (pI) were determined using the ExPASy Compute pI/Mw tool available at http://web.expasy.org/compute_pi; CBD, chitin-binding domains

Table 2. Tissue-specific expression of the two chitin synthases from *M. sexta* at different developmental stages. L-4, fourth instar larvae; L-5, fifth instar larvae; P, pupae; A, Adult. “0”, not expressed; “1”, expressed (light grey).

Tissue	/	head				midgut				abdomen		fat body				testes		ovaries	
Dev. Stage	egg	L4	L5	P	A	L4	L5	P	A	L4	L5	L4	L5	P	A	P	A	P	A
MsCHS1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0
MsCHS2	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0

Table 3. Tissue-specific expression of the chitin deacetylases from *M. sexta* at different developmental stages. L-4, fourth instar larvae; L-5, fifth instar larvae; P, pupae; A, Adult. “0”, not expressed; “1”, expressed (light grey); “2”, highly expressed (dark grey).

Tissue	/	head				midgut				abdomen		fat body				testes		ovaries	
Dev. Stage	egg	L4	L5	P	A	L4	L5	P	A	L4	L5	L4	L5	P	A	P	A	P	A
Group I MsCDA1	2	2	2	2	2	0	0	1	0	2	2	1	1	1	0	0	0	1	0
MsCDA2A	2	2	2	2	2	0	0	0	0	2	2	1	0	1	0	0	0	0	1
MsCDA2B	2	2	2	2	2	1	1	0	0	2	2	1	2	2	1	1	0	1	1
Group II MsCDA3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Group III MsCDA4	2	1	1	1	1	0	0	0	0	2	2	1	0	1	0	0	0	1	0
Group IV MsCDA5A	0	2	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
MsCDA5B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Group V MsCDA6	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0
MsCDA7	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
MsCDA8	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
MsCDA9	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0

804 **Table 4.** Tissue-specific expression of the chitinases and chitinase-like proteins from *M.*
805 *sexta* at different developmental stages. L-4, fourth instar larvae; L-5, fifth instar larvae; P,
806 pupae; A, Adult. “0”, not expressed; “1”, expressed (light grey); “2”, highly expressed
807 (dark grey).

	Tissue Dev. Stage	/	head				midgut				abdomen		fat body				testes		ovaries	
		egg	L4	L5	P	A	L4	L5	P	A	L4	L5	L4	L5	P	A	P	A	P	A
Group I	MsCHT5	1	1	0	1	0	2	0	0	0	2	1	2	1	1	0	1	0	2	0
Group II	MsCHT10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Group III	MsCHT7	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
Group IV	MsCHT8	0	0	0	0	0	0	0	1	2	0	0	0	0	0	2	0	0	0	0
Group V	MsIDGF1	1	2	2	2	2	0	0	1	1	2	2	2	2	2	2	2	2	2	2
Group VI	MsCHT6	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0
Group VII	MsCHT2	1	1	1	0	1	0	0	0	0	2	1	0	0	1	0	0	0	0	0
Group VIII	MsCHT11	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0
Group IX	MsCHT1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
Group X	MsCHT3	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Other	MsCHT-h	1	1	0	2	1	2	0	0	0	2	1	2	1	2	0	0	1	2	0

808

Figures

Fig. 1. Phylogenetic analysis of chitin deacetylases (CDA) from twelve different species; *Anopheles gambiae* (Ag), *Apis mellifera* (Am), *Bombyx mori* (Bm), *Daphnia pulex* (Dp), *Drosophila melanogaster* (Dm), *Helicoverpa armigera* (Ha), *Ixodes scapularis* (Is), *Mamestra configurata* (Mc), *Manduca sexta* (Ms), *Nasonia vitripennis* (Nv), *Tribolium castaneum* (Tc) and *Trichoplusia ni* (Tn). The accession numbers of all the proteins used are listed in Supplementary Table 1. A bootstrap analysis of 2000 replications was carried out on the trees inferred from the neighbor joining method and bootstrap values are shown at each branch of the tree. CDAs are grouped into 5 different groups: group I (pink), group II (green), group III (light blue), group IV (brown) or group V (dark purple) as described in (Dixit et al., 2008). CDAs from *M. sexta* are indicated in bold.

Fig. 2. Alignment of the catalytic domains of the chitin deacetylases containing CE4-like1 (MsCDA1 to 4) and CE4-like2 domains (MsCDA5 to 9). Alignment was performed using the ClustalW software from PBIL Expasy tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html) (Combet et al., 2000). Symbols below the aligned amino acids indicate identical (*), highly conserved (:) and conserved residues (.). Identical and highly conserved amino acids are highlighted in dark grey and light grey, respectively. The five boxes correspond to the five regions responsible for the catalytic activity described by Blair et al. (2006) and Dixit et al. (2008).

Fig. 3. Domain organization of chitinases (CHT) and chitinase-like (IDGF) proteins of *M. sexta*. For each *M. sexta* CHT from each group, one ortholog from *T. castaneum* (groups I, II, III, IV, V, VI, VII, IX, and X), *B. mori* (group h) or *D. melanogaster* (group VIII) is represented. These orthologs are representative of the typical domain organization for each group. The accession numbers of all the proteins used are listed in Supplementary Table 2. The domain organization was generated with the SMART tool (<http://smart.embl-heidelberg.de/>) by using the protein sequences. *M. sexta* proteins are represented in black while proteins from other species are in red. Catalytic GH18 domains (white boxes), chitin-binding CB14 domains (dark and red boxes), transmembrane spans (hatched boxes), polycystic kidney disease 1 PKD1 domain (white circle), and linker regions (lines) are indicated. The tree on the left is a representation of the grouping obtained by phylogenetic analysis as shown in Fig. 4.

859 (yellow), group VIII (red), group IX (grey), group X (orange) or group h (white and dotted
 860 lines). As CHT7 chitinases from Nematoda (*A. ceylanicum*, *B. malayi*, *C. elegans* and *H.*
 861 *contortus*) have a divergent domain organization from other group III chitinases (Fig. 9),
 862 they are highlighted by a hatched light blue area. CHT from *M. sexta* are indicated in
 863 bold.

864
 865 **Fig. 5.** Alignment of the 11 amino acids long motif present in GH18 catalytic domains and
 866 responsible for chitinolytic activity of chitinases for the 11 chitinases identified in *M. sexta*.
 867 The typical consensus domain described by Choi et al. (1997) with potential alternatives
 868 for four variable amino acids is shown. For each chitinase, the conserved amino acids are
 869 indicated in bold and the number of conserved amino acids regarding to the reference
 870 conserved motif is indicated in brackets. The tree on the left is a representation of the
 871 grouping obtained by phylogenetic analysis as shown in Fig. 4.

872
 873 **Fig. 6.** Alignment of the GH18 catalytic domains of the group IX chitinases from *M. sexta*,
 874 *T. castaneum*, *D. melanogaster* and *H. saltator*. Alignment was performed using the

ClustalW software from PBIL Expasy tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html) (Combet et al., 2000). Symbols below the aligned amino acids indicate identical (*), highly conserved (:) and conserved residues (.). Identical and highly conserved amino acids are highlighted in dark grey and light grey, respectively. The boxes correspond to the catalytically critical consensus sequence (position 201-208) and to highly conserved regions.

Fig. 7. Alignment of the GH18 catalytic domains of the group X chitinases from *M. sexta*, *D. plexippus*, *T. castaneum*, *C. floridanus*, *A. echinator* and *H. saltator*. Alignment was performed using the ClustalW software from PBIL Expasy tool (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html) (Combet et al., 2000). Symbols below the aligned amino acids indicate identical (*), highly conserved (:) and conserved residues (.). Identical and highly conserved amino acids are highlighted in dark grey and light grey, respectively. The box corresponds to the catalytically critical consensus sequence (position 88-93).

Fig. 8. Sequential apparition and hypothetical evolution of the eleven groups of chitinases found in *M. sexta* throughout the evolutionary tree based on fourteen species from seven different phylogenetic groups.

Fig. 9. Domain organization of group III chitinases (CHT) from *A. echinator* (AeCHT7), *A. ceylanicum* (AcCHT7), *A. gambiae* (AgCHT7), *B. malayi* (BmCHT7), *C. elegans* (CeCHT7), *D. melanogaster* (DmCHT7), *D. pulex* (DpCHT7), *H. contortus* (HcCHT7), *H. saltator* (HsCHT7), *M. sexta* (MsCHT7), *Plutella xylostella* (PxCHT7) and *T. castaneum* (TxCHT7). The accession numbers of all the proteins used are listed in Supplementary Table 2. The domain organization was generated with the SMART tool (<http://smart.embl-heidelberg.de/>) by using the protein sequences. Catalytic GH18 domains (white boxes), chitin-binding domains (dark and red boxes), transmembrane spans (hatched boxes) and linker regions (lines) are indicated.

UDP-N-acetylglucosamine