Identification and characterization of a novel chitinase-like gene cluster (AgCht5) possibly derived from tandem duplications in the African malaria mosquito, Anopheles gambiae

Jianzhen Zhang a,b, Xin Zhang b, Yasuyuki Arakane c,d, Subbaratnam Muthukrishnan c, Karl J. Kramer c, Enbo Ma a, Kun Yan Zhu b,*. 

a Research Institute of Applied Biology, Shanxi University, Taiyuan, Shanxi 030006, China
b Department of Entomology, 123 Waters Hall, Kansas State University, Manhattan, KS 66506, USA
c Department of Biochemistry, 141 Chalmers Hall, Kansas State University, Manhattan KS 66506, USA
d Division of Plant Biotechnology, College of Agriculture and Life Science, Chonnam National University, Gwangju 500-757, Republic of Korea

ARTICLE INFO

Article history:
Received 22 December 2010
Received in revised form 4 March 2011
Accepted 8 March 2011

Keywords:
African malaria mosquito
Anopheles gambiae
Chitinase
Gene cluster
Gene duplication
Glycosidase hydrolase

Insect chitinase 5 (Cht5), a well-characterized enzyme found in the molting fluid and/or integument, is classified as a group I chitinase and is usually encoded by a single gene. In this study, a Cht5 gene cluster consisting of five different chitinase-like genes (AgCht5-1, AgCht5-2, AgCht5-3, AgCht5-4 and AgCht5-5) was identified by a bioinformatics search of the genome of Anopheles gambiae. The gene models were confirmed by cloning and sequencing of the corresponding cDNAs and gene expression profiles during insect development were determined. All of these genes are found in a single cluster on chromosome 2R. Their open reading frames (ORF) range from 1227 to 1713 bp capable of encoding putative proteins ranging in size from 409 to 571 amino acids. The identities of their cDNA sequences range from 52 to 66%, and the identities of their deduced amino acid sequences range from 38 to 53%. There are four introns for AgCht5-1, two for AgCht5-2 and AgCht5-3, only one for AgCht5-4, but none for AgCht5-5 in the genome. All five chitinase-like proteins possess a catalytic domain with all of the conserved sequence motifs, but only AgCht5-1 has a chitin-binding domain. Phylogenetic analysis of these deduced proteins along with those from other insect species suggests that AgCht5-1 is orthologous to the Cht5 proteins identified in other insect species. The differences in expression patterns of these genes at different developmental stages further support that these genes may have distinct functions. Additional searching of the genomes of two other mosquito species led to the discovery of four Cht5-like genes in Aedes aegypti and three in Culex quinquefasciatus. Thus, the presence of a Cht5 gene cluster appears to be unique to mosquito species and these genes may have resulted from gene tandem duplications.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Chitinases (EC.3.2.1.14) are enzymes responsible for hydrolyzing glycosidic bonds in chitin and widely distributed in nature, including vertebrates, microorganism and even plants. Mammals are not known to synthesize chitin or metabolize chitin as a nutrient, yet the human genome encodes eight different chitinases of the glycoside hydrolase 18 (GH18) family, which play an important role in T-cell mediated inflammation and asthma (Funkhouser and Aronson, 2007; Reese et al., 2007; Shuhui et al., 2009).

In insects, chitin associates with proteins to form the cuticular exoskeleton and peritrophic matrix (PM) in the midgut lumen. During a molting cycle, a part of the old cuticle is digested while new chitin is synthesized and deposited (Reynolds and Samuels, 1996). It has been suggested that insect chitinases may have multiple functions including defense, digestion and molting (Shen and Jacobs-Lorena, 1997; Filho et al., 2002; Zheng et al., 2002; Genta et al., 2006; Zhu et al., 2008b). Indeed, a chitinase expressed in the gut of European corn borer (Ostrinia nubilalis) has been identified and it has been proposed that this enzyme is responsible for regulation of chitin content of PM and growth of O. nubilalis larvae (Khajuria et al., 2010). A chitinase was also purified from the venom gland of an endoparasitic wasp Chelonus sp. near curvimuculatus (Krishnan et al., 1994). Because of the crucial roles of chitinases in insect growth and development, these

* Corresponding author. Tel.: +1 785 532 4721; fax: +1 785 532 6232.
E-mail address: kzhu@ksu.edu (K.Y. Zhu).

0965-1748/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2011.03.001

Please cite this article in press as: Zhang, J., et al., Identification and characterization of a novel chitinase-like gene cluster (AgCht5) possibly derived from tandem duplications in the African malaria mosquito, Anopheles gambiae, Insect Biochemistry and Molecular Biology (2011), doi:10.1016/j.ibmb.2011.03.001
enzymes have been widely recognized as potential targets for developing chemical pesticides for insect control (Spindler-Barth et al., 1998; Royer et al., 2002; Hirose et al., 2010).

Insect chitinases belong to the GH18 multi-gene family with a rapid increase in the number of genes identified as the annotation of completed genome sequences of several insect species has occurred (Zhu et al., 2004, 2008a). At present, insect chitinases and chitinase-like proteins are classified into eight groups based on a phylogenetic analysis of their catalytic domains (Arakane and Muthukrishnan, 2010). Among these chitinases and chitinase-like proteins, chitinase 5 (Cht5) is classified into group I.

To date, only a single Cht5 gene has been identified from each of several insect genomes and the representatives of this gene have been well characterized in several lepidopteran and coleopteran species (Kim et al., 1998; Shinoda et al., 2001; Zheng et al., 2002; Ahmad et al., 2003; Fitches et al., 2004; Bolognesi et al., 2005; Zhu et al., 2008a). All insect Cht5s have a typical multidomain structural organization that includes a signal peptide, a catalytic domain, a PEST-like linker region enriched in proline (P), glutamic acid (E), serine (S) and threonine (T) that is heavily glycosylated, and a cysteine-rich chitin-binding domain. The sizes of Cht5 enzymes range from 552 to 586 amino acid residues. The transcripts of Cht5 are mainly detected in the epidermis and the gut, and its expression increases during the molting process (Royer et al., 2002). In Helicoverpa armigera, Cht5 is also expressed in the fatbodies (Ahmad et al., 2003). However, the expression of Cht5 has not been reported in haemocytes from any insect species. In Manduca sexta, the transcript level of Cht5 can be induced by 20-hydroxyecdysone, but it is suppressed by the juvenile hormone analog, fenoxycarb (Kramer et al., 1993). The recombinant protein expressed in an insect cell line showed high levels of chitinolytic activity (Gopalakrishnan et al., 1995; Zheng et al., 2003; Ahmad et al., 2003).

It is likely that Cht5 may be involved in chitin turnover associated with molting. In Tribolium castaneum, RNA interference was performed to silence TcCh5. The insects that were injected with double-stranded RNA (dsRNA) for TcCh5 exhibited a lethal phenotype only at the pharate adult stage. At the time of death, some of the adult cuticle was visible under the old pupal cuticle which was not shed, suggesting that TcCh5 is required for pupal–adult molting (Zhu et al., 2008b). All these results indicate that Cht5 is an essential enzyme for insect growth and development.

Although insects have been known to have only a single Cht5 gene, our recent studies have revealed a novel Cht5 gene cluster consisting of multiple chitinase-like genes in three mosquito species. In this paper, we report: 1) identification of a cluster of five An. gambiae Cht5-like genes (AgCht5-1, AgCht5-2, AgCht5-3, AgCht5-4 and AgCht5-5) and their chromosomal localization, 2) characterizations of their gene models and developmental expression patterns, and 3) the results of a comparative investigation on Cht5 gene clusters in two other mosquito species including Aedes aegypti and Culex quinquefasciatus. This is the first demonstration of gene duplication of this group of chitinase genes, which may be unique to the mosquito lineage.

2. Materials and methods

2.1. Insect culture

An. gambiae was obtained from the Malaria Research and Reference Reagent Resource Center (MR4, Manassas, VA) and has been cultured in the Department of Entomology at Kansas State University, Manhattan, KS since 2005. The colony was maintained as described previously (Zhang and Zhu, 2006)
RT-PCR was carried out in a 25-μl reaction mixture containing 1 μl template cDNA, 12.5 μl Taq Master Mix (Fermentas), 0.2 μl of each primer and sterilized water. The thermal cycle program for RT-PCR consisted of an initial denaturation at 94 °C for 1 min followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 45 s, and a final extension at 72 °C for 5 min. PCR products were analyzed on a 2% agarose gel. Three biological replications (i.e., 3 independent preparations of total RNA), each with three repeated PCR runs, were performed in this analysis.

3. Results

3.1. Full-length cDNAs and the deduced amino acid sequences of five AgCht5 genes

The conceptual translation of the An. gambiae gene model XP_001237469.2 predicts a 2095 amino acid-long protein with five catalytic domains. However, analysis of the sequences of five cDNA clones (EST clone numbers: 19600449629438 for AgCht5-1, 19600449653107 for AgCht5-2, 19600449656904 for AgCht5-3, NAP1-P158-B-06-5 for AgCht5-4, and 19600449684410 for AgCht5-5) that were obtained from the MR4 failed to provide evidence for a long transcript that bridges genomic sequences encoding adjacent chitinase catalytic domains predicted by this gene model. Instead, we could detect sequences that were presumed to be introns in this gene model at the 5'-ends or 3'-ends of the five full-length chitinase cDNA sequences that we have characterized.

The additional sequences at the 3'-ends of these clones (which were not included in the gene model XP_001237469.2) had stop codons and polyadenylation signal sequences and short poly A tails. The additional sequences at the 5'-ends of the full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions. The sequence data we obtained from full-length cDNA clones included start codons followed by signal peptide coding regions.

Fig. 1. Genome structure, exon/intron organizations and protein domain architectures of five AgCht5 genes. A) Genome structure of AgCht5-1, AgCht5-2, AgCht5-3, AgCht5-4 and AgCht5-5. The light brown box and blue line represent DNA sequences of each gene and the linker region, respectively, between the ORFs. The length of linker region is marked under the blue line and by a yellow triangle. B) Exon and intron organizations: Exons are shown by green boxes whereas introns are shown by pink lines. C) Domain architectures of predicted AgCht5 proteins: Predicted signal peptide, catalytic domain and chitin-binding domain are boxed in yellow, gray and purple, respectively, whereas linker regions are shown by blue lines.

patterns were further examined in eggs collected at 12, 24, 36, 48 h and fourth-instar larvae with trace amounts of transcripts detected in the eggs and fi rst-instar larvae. Their detailed expression patterns when we used unique primer sets designed to amplify cDNAs from specific regions of AgCht5. This finding prompted us to hypothesize that AgCht5 actually is a gene cluster consisting of multiple genes and the different catalytic domains might be encoded by different genes. To address this question, we utilized the cDNA sequence of each domain of AgCht5 to search the An. gambiae genome, a multiple-member Cht5 gene cluster in insects.

4. Discussion

Currently, the deduced Cht5 proteins from different insect species are grouped into one clade in the phylogenetic analysis of insect chitinases (Arakane and Muthukrishnan, 2010). All chitinases possess a typical multiple domain structural organization consisting of a signal peptide, an N-terminal catalytic domain with four conserved motif sequences, KXXXXXGGW, FDGXDLDWEYP, C-(X11)-C-(X5)-C-(X9)-C-(X12)-C-(X7)-C spacing of six residues expected of chitinases (Fig. 2). As supported by the sequences expected of chitinases (Fig. 2), AgCht5-1 and AgCht5-3 were apparently expressed during all the pupal stage and exhibited similar expression patterns, whereas AgCht5-2 and AgCht5-5 were mainly expressed in 0- and 10-h pupae. However, the expression of AgCht5-4 appeared to gradually increase with pupal development from 0 to 34 h.

3.2. Phylogenetic analysis of five deduced AgCht5 protein sequences

To explore the relationship among the insect Cht5s, a phylogenetic tree was constructed based on the sequences of their catalytic domains. Results showed that all of the insect Cht5s fall into two branches supported by a bootstrap value of 100 after 5000 replicates (Fig. 4). The first group represents the mosquito Cht-1 and all other well characterized insect Cht5s with chitinase activities, whereas the second group represents the remaining mosquito Cht5s. Apparently, AgCht5-1, AaCht5-1 and CqCht5-1 from the three mosquito species are more closely related and might represent mosquito orthologs of insect Cht5 enzymes. In contrast, the mosquito Cht5-2, Cht5-3, Cht5-4, and Cht5-5 are clustered in another branch that may be encoded by genes derived from an ancestral Cht5-1 by gene duplications. Cht5-2s from three mosquito species close together with robust bootstrap value, suggest they are also orthologs.

3.3. Developmental stage-dependent expression patterns of five AgCht5 genes

The developmental stage-dependent expression patterns of different AgCht5 transcripts were determined by RT-PCR. The levels of transcripts of the five AgCht5 genes were apparently higher in third- and fourth-instar larvae (Fig. 5A). Four genes including AgCht5-1, AgCht5-2, AgCht5-3 and AgCht5-5 were expressed at all developmental stages, whereas AgCht5-4 was expressed mainly in third- and fourth-instar larvae with trace amounts of transcripts detected in the eggs and first-instar larvae. Their detailed expression patterns were further examined in eggs collected at 12, 24, 36, 48 and 60 h after deposition by blood-fed females (Fig. 5B). High transcript levels were detected in 36-h eggs for all of the five AgCht5s. However, no detectable expression was found for AgCht5-2 in 60-h eggs and for AgCht5-5 in 12-h eggs. On the other hand, AgCht5-4 was scarcely detected in mature eggs. Similarly, the expression patterns of the five different AgCht5 genes were also examined in pupae collected at 0, 10, 20, 30 and 34 h after pupation (Fig. 5C). AgCht5-1 and AgCht5-3 were apparently expressed during all the pupal stage and exhibited similar expression patterns, whereas AgCht5-2 and AgCht5-5 were mainly expressed in 0- and 10-h pupae. However, the expression of AgCht5-4 appeared to gradually increase with pupal development from 0 to 34 h.

Table 2

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amino acid residue</th>
<th>Catalytic domain</th>
<th>Chitin-binding domain</th>
<th>Availability of ESTa</th>
<th>Genome location</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgCht5-1</td>
<td>571</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>chr2R:21,584,333 – 21,587,318</td>
</tr>
<tr>
<td>AgCht5-2</td>
<td>412</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>chr2R:21,582,374 – 21,583,826</td>
</tr>
<tr>
<td>AgCht5-3</td>
<td>413</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>chr2R:21,578,829 – 21,580,211</td>
</tr>
<tr>
<td>AgCht5-4</td>
<td>409</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>chr2R:21,576,773 – 21,578,085</td>
</tr>
<tr>
<td>AgCht5-5</td>
<td>446</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>chr2R:21,573,544 – 21,574,884</td>
</tr>
</tbody>
</table>

a Based on the Anopheles gambiae EST database from NCBI.

the names AaCht5-1, AaCht5-2, AaCht5-3 and AaCht5-4 for those identified in Ae. aegypti, and CqCht5-1, CqCht5-2 and CqCht3 for those identified in C. quinquefasciatus. Analysis of their domain architectures indicated that all the deduced Cht5 proteins from the three mosquito species have a catalytic domain, but only the first Cht5 protein in each species (i.e., AgCht5-1, An. gambiae 1-and CqCht5-1) contains a chitin-binding domain (Fig. 3).
Our further studies confirmed a similar clustering of $Cht5$ genes in other mosquito species. Based on our genome search, we identified four $Cht5$-like genes in $Ae. aegypti$ and three in $C. quinquefasciatus$. By aligning these mosquito $Cht5$ proteins with other known insect $Cht5$ proteins, we found that only one $Cht5$ catalytic domain ($AgCht5-1$, $AaCht5-1$ or $CqCht5-1$) from each mosquito species was clustered with other domains from known insect $Cht5$ proteins with a bootstrap value of 100 (Fig. 4), whereas the remaining $Cht5$s from these three mosquito species were grouped into a different cluster. These results suggest that $AgCht5-1$,
AaCht5-1 and CqCht5-1 are predicted to be catalytically active, catalytic mechanism (Watanabe et al., 1994). Thus, AgCht5-1, lytically active because it probably serves as a proton donor in the enzyme. The third residue E is crucial for a chitinase being catalytically active because it is located in or near the catalytic site of the enzyme. The presence of the third residue E suggests that this gene cluster could be evolved from gene duplications. Gene duplication events are generally considered to be essential in the evolution of gene families, which facilitate the amplification of chitinase genes in Drosophila melanogaster and T. castaneum suggests that amplification of this subgroup of chitinase genes is of recent origin.

The catalytic and chitin-binding domains are two important structural components of chitinases. Sequence motif analysis showed that all mosquito Chs5 proteins possess a catalytic domain, but only three of the proteins, AgCht5-1, AaCht5-1 and CqCht5-1, contain the signature sequence DWEPYF within conserved region II, which is known to be located in or near the catalytic site of the enzyme. The third residue E is crucial for a chitinase being catalytically active because it probably serves as a proton donor in the catalytic mechanism (Watanabe et al., 1994). Thus, AgCht5-1, AaCht5-1 and CqCht5-1 are predicted to be catalytically active, which are similar with all the other known insect Chs5s. In contrast, all of the other predicted Chs5-like proteins from the three mosquito species are likely to be catalytically inactive because the E residue is replaced by L in these proteins. These proteins may have carbohydrate-binding capability.

**Fig. 4.** Phylogenetic analysis of catalytic domain sequences of putative chitinase 5 proteins from different species including Aedes aegypti (Aa, XP_001656234.1 for AaCht5-1, XP_001656233.1 for AaCht5-2, XP_001656232.1 for AaCht5-3, and XP_001656231.1 for AaCht5-4), Anopheles gambiae (Ag, HQ456129 for AgCht5-1, HQ456130 for AgCht5-2, HQ456131 for AgCht5-3, HQ456132 for AgCht5-4, and HQ456133 for AgCht5-5), Bombus mandarina (Bma, AAB47538), Choristoneura fumiferana (Chf, AAM43792), Culex quinquefasciatus (Cq, XP_001863384.1 for CqCht5-1, XP_001863385.1 for CqCht5-2, and XP_001863386.1 for CqCht5-3), Drosophila melanogaster (Dm, CG9307), Helicoverpa armigera (Ha, AAQ1786), Hyphantria cunea (Hc, AAB47537), Lacanobia oleracea (Lo, CAFO5663), Manduca sexta (Ms, P36362), Spodoptera frugiperda (Sf, AAS18266), Spodoptera litura (Sl, A0023107), and Tribolium castaneum (Tc, AY675073). The phylogenetic tree was constructed using Mega 4 software (Tamura et al., 2007). Bootstrap values are obtained by the neighbor-joining method using 5000 replications. Bootstrap values are indicated only when greater than 40%.

**Fig. 5.** The expression patterns of five AgCht5 genes in Anopheles gambiae as evaluated using RT-PCR. A) Gene expression patterns in eggs (EG), first- (L1), second- (L2), third- (L3), fourth- (L4) and fifth-instar larvae (L5); and adults (AD). B) Gene expression patterns in 12-, 24-, 36-, 48- and 60 h-old eggs as shown by EG12, EG24, EG36, EG48, and EG60, respectively. C) Gene expression patterns in 0-, 10-, 20-, 30- and 34-h-old eggs as shown by PU00, PU10, PU20, PU30 and PU34, respectively. AgRps3 was used as reference gene for RT-PCR analysis.
generation of new genes with new functions. Gene duplication can occur via three major mechanisms: segmental duplication (of the whole genome, of one to a few chromosomes or of large parts of a chromosome), tandem duplications (of one to a few adjacent genes), and retroposition or other transposition events (Kong et al., 2007). Among these, tandem and segmental duplication events contribute mostly to the generation of new members in nuclear gene families. Tandem duplicates are copies of a nearby gene that are within short intron distances of each other and may harbor some interesting biology. Gene expansion by tandem duplication is common in cytoplasmic P450 gene evolution (Ai et al., 2010; Baldwin et al., 2009). Five AgCht5 genes are clustered together in chromosome 2R with no other intervening genes. The minimum distance is 340 bp between two chitinase ORFs and the maximum is 2045 bp, suggesting that these genes may be derived from tandem duplications. Gene duplication and loss according to a birth–death model of evolution is a feature of the evolutionary history of the family 18 (GH18) of chitinases (Funkhouser and Aronson, 2007). Based on above information, we propose that AgCht5 gene cluster may evolve primarily from tandem duplication.

Because the transcripts of all the five AgCht5 genes can be detected at various developmental stages in An. gambiae (Fig. 5), these genes are transcriptionally and appear to be independently regulated. However, the deduced AgCht2-2, AgCht5-3, AgCht5-4 and AgCht5-5 proteins lack an essential catalytic glutamic acid residue for hydrolytic activity (Watanabe et al., 1994; Lu et al., 2002, Zhang et al., 2002, Zhu et al., 2008a,b). Therefore, they are presumed to act as carbohydrate-binding proteins or lectins rather than as enzymes because they may not have any catalytic activity (Goormachtig et al., 2001). Nevertheless, it should be pointed out that a chitinase-like protein without catalytic activity may still play an important role in insect development. For example, insect imaginal disc growth factors (IDGFs) are chitinase-like proteins that are structurally related to chitinases but do not possess enzymatic activity. However, TcIDGF4 identified in T. castaneum might be involved in cell proliferation and contributed to adults ecysis (Zhu et al., 2008b). Further studies will be necessary to elucidate the biological function of each of the five duplicated Cht5 genes in An. gambiae.

Acknowledgements

Authors thank Sharon R. Starkey for maintaining mosquito colonies. This research was supported in part by Kansas Agricultural Experiment Station and K-State Arthropod Genomics Center funded by K-State Targeted Excellence program at Kansas State University, Experiment Station and K-State Arthropod Genomics Center funded An. gambiae in Manhattan, Kansas.

References


J. Zhang et al. / Insect Biochemistry and Molecular Biology xxx (2011) 1–8

Please cite this article in press as: Zhang, J., et al., Identification and characterization of a novel chitinase-like gene cluster (AgCht5) possibly derived from tandem duplications in the African malaria mosquito, Anopheles gambiae, Insect Biochemistry and Molecular Biology (2011), doi:10.1016/j.ibmb.2011.03.001