

Attacin gene sequence variations in different ecoraces of tasar silkworm *Antheraea mylitta*

Rati Sudha, Geetha N. Murthy, Arvind K. Awasthi, Kangayam M. Ponnuvel*

Seribiotech Research Laboratory, Carmelram Post, Kodathi, Bangalore - 560 035, Karnataka, India. Phone: +91 8050614267, +91-8903891130, Kangayam M. Ponnuvel - Email: kmpvel@yahoo.com; *Corresponding author

Received October 17, 2015; Accepted October 19, 2015; Published October 31, 2015

Abstract:

Attacin gene exists as paralogous conversion and is being used for identification of strain variations in insects based on the sequence variation. Hence, a study was undertaken to analyze the sequence variation of the attacin gene isoforms in the tasar silkworm *Antheraea mylitta* that exists in the form of different ecoraces depending upon the environment, food plant and location. Comparison of the previously reported attacin sequences with the DNA sequences of attacin A and B genes revealed six amino acid substitutions among the sequences of the ecoraces which however did not affect the functional domain of Attacin. The generated dendrogram clearly indicated unique branches for each ecorace with two separate gene clusters for attacin A and B. The Sarihan ecorace formed a separate sub-group under both the gene clusters. The present study also revealed the presence of Attacin_N Superfamily domain exclusively in Exon I separated from the Attacin_C Superfamily domain that was present in Exon II and part of Exon III, a prominent character of attacin gene. The phylogenetic reconstruction analysis of attacin gene in *A. mylitta* supported the common evolutionary origin of attacin genes belonging to the Lepidoteran and Dipteran families that formed two separate clusters.

Background:

Insects have an efficient humoral immune system against microorganisms, which attributes to their evolutionary success in occupying almost all habitats in nature. They mainly rely on innate immunity as defense mechanism [1]. The immune status of insects is defined by a set of proteins that is absent in the naive state [2]. In this defence system, antibacterial proteins play a major role in eliminating invading bacteria. Boman *et al.* [3] observed that, injection of bacteria into pupae of *Hyalophora cecropia* moths resulted in the synthesis of immune proteins. These proteins were purified and characterized as novel classes of antibacterial proteins called cecropins and attacins.

In Lepidopteran insects, attacin, cecropin and lysozymes are the major antibacterial proteins secreted into the hemolymph upon infection with bacteria as shown by Boman *et al.* (1991) [4]. In mulberry silkworm, five types of anti-bacterial proteins viz. cecropin, attacin, lebocin, lysozymes and moricin have been identified and three subtypes of cecropin (A, B and D) have been reported. Among the anti-bacterial proteins, attacin is highly expressed and available abundantly. Attacins are glycine-rich immune proteins and the most widespread among the different antibacterial proteins in insect species. They were first reported in the moth *H. cecropia* in response to bacterial infection [5]. These attacins are rather heat-stable, with 40% of

the antibacterial activity remaining even after 1 h at 100°C and are immune effectors molecules that can inhibit the growth of gram negative bacteria. The attacins are synthesized as pre / pro-mature peptides with the mature peptide typically being about 190 amino acids in length and forming a 'random coil' structure in solution [6]. There are six isoforms of attacins viz. A, B, C, D, E and F and are divided into two groups according to their amino acid composition and amino-terminal sequences. Attacins A-D constitutes the basic group and attacins E and F the acidic group [7]. The attacins A-D have slightly higher levels of threonine, glutamic acid, lysine and tryptophan. The molecular weights of attacins range from 20-23 kDa. In the *Drosophila* genome, a family of four genes encodes attacins A, B, C and D. Attacin C is similar to A and B, while, D is more divergent [8].

In India, tropical tasar cocoons yield one of the forms of silk spun by tasar silkworm *Antheraea mylitta* that are polyphagous feeding on leaves of a variety of food plants like Arjun (*Termanalia arjuna*), Asan (*T. tomentosa*) and Sal (*Shorea robusta*) found in the forests of eight states of India namely Jharkhand, Chhattisgarh, Orissa, Andhra Pradesh, Maharashtra, West Bengal and Uttar Pradesh. Since, *A. mylitta* is widely distributed in nature, various populations have become geographically isolated over centuries and have adapted to

those particular ecological niches. These populations are referred to as ecological races or eco-races. Hence, it is reasonable to expect intra and inter species sequence variation in the different ecoraces of *A. mylitta* as they form ecoraces depending upon the environment, food plants and locations.

Attacin is most effectively expressed as isoforms following bacterial infection that tend to become different paralogous conversions depending upon the insect species / strain [5, 8]. These isoforms are typically organized in small clusters that appear to be in a dynamically steady state where, new genes are continuously produced by gene duplication, while, others are lost due to mutation [9]. Natural genetic variation surveyed in alleles of Attacins A, B, and C recovered from a wild population of North American *D. melanogaster* revealed that, the overall level of nucleotide diversity is quite high without excess of amino acid polymorphism. Attacins A and B have experienced multiple paralogous gene conversion events and a recent conversion has created a novel haplotype that subsequently increased rapidly in frequency. A number of polymorphisms observed, including a null allele of Attacin A that may affect the functional capacity of the immune response [9]. In *D. melanogaster* strain variation was identified based on the sequence variation in the attacin gene sequences. Hence, as the tasar silkworm *A. mylitta* exists as different ecoraces depending upon the environment, food plants and locations, a study was undertaken to analyze the sequence variation of the attacin gene isoforms in the eco races.

Methodology:

Selection of tasar silkworm ecoraces

Wild cocoons were collected from East and West Singhbhum, Ghatshila, Hazaribagh and Santhal Pargana districts of the State of Jharkhand in India. Six ecoraces of *A. mylitta* viz. Daba, Laria, Modia, Sarihan and Jatta Daba, and Andhra Local were selected and genomic DNA extracted using standard phenol-chloroform extraction [10]. Thirty samples (six / eco-race) of genomic DNA were screened from the above six ecoraces.

Designing of primers

The attacin gene sequence of *A. mylitta* was retrieved from NCBI databases having accession numbers DQ666489 (Attacin A) and DQ666490 (Attacin B). To identify the sequence variation in the attacin gene, two sets of primers were designed. The forward primer binding site for attacin A was at 45bp with 5'- TATTCCTCGIGTCCGTCCTT-3'; The reverse primer binding site was at 646bp with 5'- CCAGCCGGCATTAAAA TDTA -3'; The forward primer for attacin B binding site was at 76bp with 5'- CACCGACTTCATCAACCGTA-3'; The reverse primer was at binding site at 399bp with (5'- TGGTTTATTGTA AAATTACGCTAGTTA -3')

Polymerase chain reaction (PCR) and product analysis

PCR was done in an MJ Research thermal cycler, PTC200 (MJ Research Inc., 149 Grove Street, Watertown, Massachusetts, USA) using 20µl reaction containing 4µl 5x PCR buffer, 0.5µl 10mM dNTPs, 2.0µl 25 mM MgCl₂, 10 pmoles of the forward and reverse primers each and 0.2µl of TaqDNA polymerase (MBIFermentas) and remaining distilled water. The PCR schedule was 94° C for 2 min followed by 30 cycles of 94° C for 30s, 55° C for 30s, 72° C for 2 min and final extension of 10 min at 72° C. The PCR was resolved on 1.5% agarose gel in Tris-boric acid/ EDTA buffer with a constant voltage of 80 in parallel with standard markers. The molecular weight of PCR

product was calculated using the software program viz. "INCHWORM" (www.molecularworkshop.com/pl/inch2001.pl).

In order to sequence the PCR amplified products of the Attacin A and B genes, the respective genes were cloned in the TA cloning vector (pTZ57R/T, Fermentas Pvt. Ltd.). Further, to confirm the presence of the insert in the vector, PCR amplification was carried out with the gene specific and M13 primers using plasmid DNA as template. The PCR product analysis indicated amplification with both types of primers. The results indicated that, both the Attacin genes were cloned in the TA cloning vector. These plasmid DNA samples were further purified through DNA purification column (Promega) for sequencing with M13 primers at MWG Pvt. Ltd., Bangalore. Further the sequence data is submitted to GenBank under the following accession numbers KT587335, KT587336, KT587337, KT587338, KT587339, and KT587340.

Spidey

The nucleic acid sequences were analyzed for the presence of exons and introns by Spidey program (a tool for mRNA and genomic alignments) [11].

Multiple sequence alignment and phylogenetic analysis

The nucleotide sequences of antibacterial gene in different silkworm races were compared using the multiple alignments Clustal W program [12]. The result output was obtained as a dendrogram with bootstrap values with 1,000 replications. The full-length amino acid sequence of attacin gene was compared with insect antibacterial genes in databases using BLAST (BLASTp) [13]. The phylogenetic relationship of attacin gene in the Lepidoptera order was analyzed using the MEGA 3.1 program [14]. The tested Lepidoptera comprised of the following families: Bombycidae, Lymantriidae, Saturniidae and Spingidae.

Results and Discussions:

Amplification of ecoraces with designed primers

Gandhe *et al* (2006) [7] constructed a cDNA library from fat body tissues of *E.coli* challenged *A.mylitta* larvae and randomly sequenced a large number of EST clones from the library. The different ESTs were classified into categories such as immune related, housekeeping, hypothetical insect proteins and hypothetical non-insect proteins. Of the 1412 ESTs, 432 (31%) showed similarity to known insect immune proteins. Among the 432 clones, 96 (22%) ESTs belonged to attacin genes that were in the form of isoforms. The sequences of these isoforms were retrieved from the databases with accession numbers DQ666489 (bp = 1628) and DQ666490 (bp = 365), respectively, and the forward and reverse primers were designed in the respective attacin gene sequence and were used to amplify the genomic DNA of five ecoraces of *A. mylitta* viz. Laria, JattaDaba, Daba Natural, Modia, Sarihan and Andhra Local (six individuals / ecorace). The information on primer binding site, amplification product size and the primer sequences are given in **Table 1 (Supplementary Material)**. The primers of Attacin A gene produced 1490 bp and Attacin B produced 365 bp long fragments. Primers designed for attacin A and B genes amplified samples of all the five ecoraces (**Figure 1**).

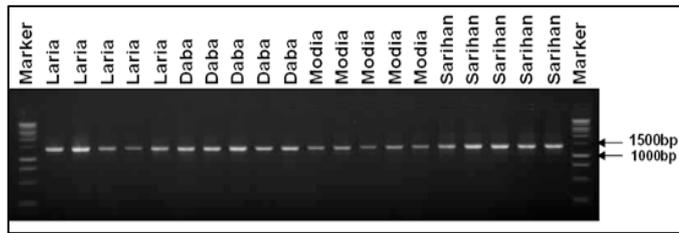


Figure 1: Amplification of attacin gene with gene specific primers

Comparison of attacin gene sequences on the natural genetic variations

A survey on the natural genetic variations in alleles of attacin A and B recovered from the five selected *A. mylitta* ecoraces and comparison of the said two attacin gene sequences indicated six mis-matches among the sequences of the ecoraces (Figure 2). Under attacin A cluster, the ecoraces Laria and Modia formed one sub-group, while, Jatta Daba, Daba and Sarihan formed three separate sub-groups. Similarly, under attacin B cluster, the ecoraces Jatta Daba and Daba formed one sub-group, while; Laria, Modia and Sarihan formed three separate sub-groups (Figure 3).

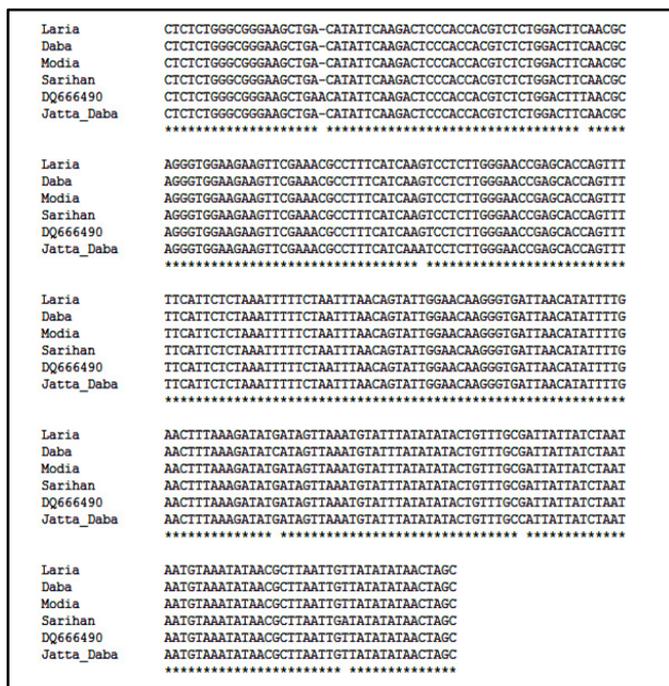


Figure 2: Attacin B gene sequence from different races of Tasar silkworm

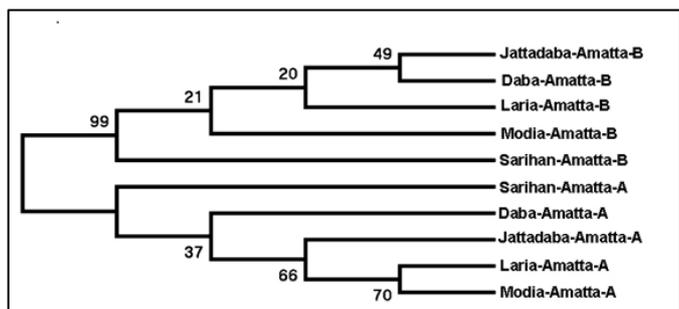


Figure 3: Dendrogram of attacin gene from different *A. mylitta* ecoraces

Similar results were obtained by Wang *et al.* (2008) [15] on comparison of previously reported attacin cDNA sequences with the DNA sequence of attacin A and B genes in *Glossina morsitans*, wherein two nucleotide substitutions were observed at positions 127 and 504. Further, they also observed that, alignment of cDNAs with those of the published cDNA sequences showed that, none of them contained the polymorphisms associated with positions 127 and 504. Thus, it is difficult to point out that, the variations may be due to the presence of other attacin coding loci in the genome or due to allelic variations that may have given rise to the polymorphisms observed.

Comparison of the previously reported attacin sequences with the DNA sequences of attacin A and B genes revealed, nucleotide variations at 20 different amino acid positions (Table 2 - Supplementary material). At amino acid position 21, insertion of nucleotide was observed, while at other positions, substitutions were observed. Among these substitutions, six revealed amino acid substitutions viz. Glutamic acid to Glutamine, Serine to Asparagine, Valine to Leucine, Leucine to Phenylalanine, Methionine to Asparagine and Serine to Asparagine.

Earlier reports have indicated that, attacin genes reveal genetic variation in the promoter and coding regions of *D. melanogaster* [9]. They also reported that, attacin gene evolved from a common ancestor and conserved among different insect taxa and that, the overall level of nucleotide diversity is prominent in attacin A and B genes, but, there is no excess of amino acid polymorphism.

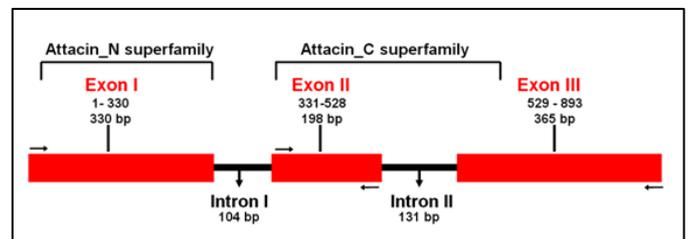


Figure 4: Genomic organization of attacin gene of *A. mylitta*

Analysis of genomic organization of attacin gene

The full length attacin A and B gene sequences obtained from the genomic DNA of Laria ecorace were analyzed and the exons and introns were predicted using Spidey Program (Figure 4). The results indicated that, each attacin gene sequence had three exons with two introns. The lengths of the introns were 104bp and 131bp, respectively. Among the three exons, Exon III had the maximum length of 365bp followed by Exon I and II, 330bp and 190bp, respectively. The attacin domains analyzed in the predicted Exons revealed the presence of Attacin_N Superfamily domain exclusively in Exon I. The Attacin_C Superfamily domain was observed in Exon II and part of Exon III.

Sun *et al.* (1991) [16] reported that, only two functional attacin genes were revealed in their studies on *H.cerropia*, one corresponding to the acidic and one to the basic attacin cDNA. The two attacin genes are transcribed in opposite directions and interrupted at homologous positions by two introns. In the present study also, each attacin gene sequence had three exons with two introns as reported by Sun *et al.* (1991) [16]. Attacins,

sarcotoxin II and dipterin share certain protein domains, which could be of functional importance. The proposed functional domains coincided with the exons (exons 2 and 3) of the attacins. Since, a certain degree of sequence similarity exists between exons 2 and 3 of the attacins, it is possible that, they appeared through duplication of one ancestral exon. This is consistent with the fact that, dipterin contains only one such protein domain, while, attacin and sarcotoxin II contain two copies of the domain [17]. In order to confirm the functional activity of attacin genes in different eco races, the translated protein sequence was compared with attacin protein sequence.

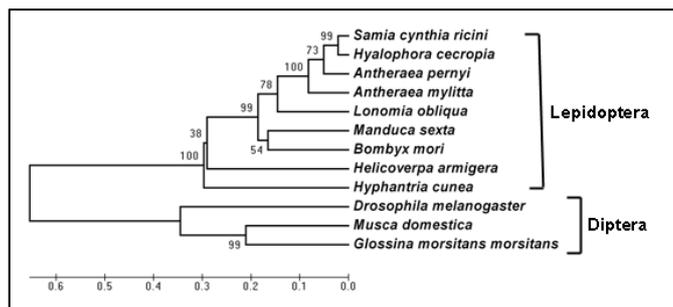


Figure 5: Phylogenetic tree of attacin gene from different insect species

Phylogenetic reconstruction analysis of attacin gene

The phylogenetic reconstruction analysis of attacin gene in *A. mylitta* supported the common evolutionary origin of attacin genes belonging to the Lepidoteran and Dipteran families, which formed two separate clusters. In the Lepidopteran cluster, all members of the order Saturniidae formed separate sub-cluster. In the same cluster, however, *Bombyx mori* of order Bombycidae, formed separate sub-cluster with *Manduca sexta* of order Sphingidae [Figure 5]. The sequence of attacin genes between the two families showed significant similarity. Further, the presence of RVRRL (43-46 AA) cleavage site between the P and G domains suggested that, processing of Attacin in *A. mylitta* is as observed in other insects.

Conclusion:

Primers designed for attacin A and B genes amplified in all the six ecoraces viz. Laria, Jatta Daba, Daba Natural, Modia Sarihan and Andhra Local confirming the presence of both the genes in their genome. Survey on the natural genetic variations in alleles of attacin A and B from the five different *A. mylitta* ecoraces through comparison of their gene sequences indicated prominent overall level of nucleotide diversity without excess of amino acid polymorphism. The six amino acid substitutions observed did not affect the functional domain of Attacin. The dendrogram clearly indicated unique branches for each ecorace and separation of clusters for attacin A gene from that of attacin

B gene. The Sarihan ecorace formed a separate sub-group under both the gene clusters. The excessive polymorphism at nucleotide level and no change in the amino acid sequence indicates that, the attacin gene is flexible in accepting changes in the nucleotide. However, amino acid sequences are not excessively polymorphic. In the present study also, results indicated single nucleotide polymorphism in attacin gene, which could be useful to distinguish different ecoraces of tasar silkworm, *A. mylitta*. The attacin domains analyzed in the predicted Exons revealed the separation of the two domains viz. Attacin_N Superfamily domain exclusively in Exon I, while, the Attacin_C Superfamily was observed in Exon II and part of Exon III. This separation of the above two domains through exons which is a prominent character of attacin gene, was also obtained in the present study.

Acknowledgements:

This research work was supported with grants from Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/PR 4925/PBD/19/111/2004).

References:

- [1] Kishimoto K *et al. Insect Biochem. Mol. Biol.* 2002 **2**:881 [PMID: 12110295]
- [2] Kang D *et al. Insect BiochemMol Biol.* 1996 **26**:177 [PMID: 8882660]
- [3] Boman HG *et al. Mem Inst Oswaldo Cruz.* 1987 **82**:115 [PMID: 3334401]
- [4] Boman HG *et al. Eur. J. Biochem.* 1991 **201**: 849 [PMID: 1915368]
- [5] Hultmark D *et al. Embo J.* 1983 **2**:571 [PMID: 6628360]
- [6] Gunne H *et al. Eur. J. Biochem.* 1990 **187**:699 [PMID: 2406140]
- [7] Gandhe AS *et al. BMC Genomic* 2006 **7**:184 [PMID: 16857061]
- [8] Hedengren Met *al. Biochem. Biophys. Res. Commun.* 2000 **279**:574 [PMID: 11118328]
- [9] Lazzaro BP *et al. Genetics* 2001 **159**:659 [PMID: 11606542]
- [10] Reddy KD *et al. Genome* 1999 **42**:1057 [PMID: 10659770]
- [11] Wheelan SI *et al. Genome research* 2001 **11**:1952 [PMID: 11691860]
- [12] Thompson JD *et al. Nucleic Acids Research* 1994 **24**:4876 [PMID: 9396791]
- [13] Altschul MD *et al. Nucleic Acids. Res.* 1997 **25**: 3389 [PMID: 9254694]
- [14] Kumar S *et al. Bioinformatics* 2004 **5**:150 [PMID: 15260895]
- [15] Wang J *et al. Insect Molecular Biology* 2008 **17**:293 [PMID: 18477243]
- [16] Sun S *et al. Eur. J. Biochem.* 1991 **196**: 247 [PMID: 2001705]
- [17] Wicker C *et al. J. Biol. Chem.* 1990 **265**:22493 [PMID: 2125051]

Edited by P Kanguane

Citation: Sudhaet *al. Bioinformation* 11(10): 481-485 (2015)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.

Supplementary material

Table 1: Details of Attacin gene primers

Gene	Clone gene	Primer	Primer sequence	PCR Product Size (bp)
Attacin A	DQ666489	Forward	TATTCCTCGTGTCCGTCCTT	1628
		Reverse	CCAGCCGGCATTAAAATDTA	
Attacin B	DQ666490	Forward	CACCGACTTCATCAACCGTA	365
		Reverse	TGGTTTATTGTAAAATTACGCTAGTTA	

Table 2: Codon changes between two attacin genes of *A. mylitta*

S. No.	Residue position	Amattacin A	Amattacin B	Amino acid
1.	21	AGA	---	Arginine
2.	28	TAT	TAC	Tyrosine
3.	29	ATT	ATC	Isoleucine
4.	31	GAG	CAA	Glutamic acid→Glutamine
5.	41	AGT	AAC	Serine→Asparagine
6.	64	AAG	AAA	Lysine
7.	69	GGG	GGA	Glycine
8.	77	GCA	GCG	Alanine
9.	78	ATT	ATC	Isoleucine
10.	81	GTT	CTT	Valine→Leucine
11.	82	GAT	GAC	Aspartic acid
12.	83	CTT	TTC	leucine→Phenylalanine
13.	87	ATG	AAT	Methionine→ Asparagine
14.	89	CTG	TTG	Leucine
15.	91	GCT	GCA	Alanine
16.	94	GCC	GCT	Alanine
17.	102	ACA	ACT	Threonine
18.	129	TTG	CTG	Leucine
19.	159	AAC	AAT	Asparagine
20.	206	AGC	AAC	Serine→Asparagine