Published in final edited form as:

Mol Phylogenet Evol. 2014 July ; 76: 127–133. doi:10.1016/j.ympev.2014.03.014.

Evolutionary Diversification of Aminopeptidase N in Lepidoptera by Conserved Clade-specific Amino Acid Residues

Austin L. Hughes

Department of Biological Sciences, University of South Carolina, Columbia SC 29208 USA

Abstract

Members of the aminopeptidase N (APN) gene family of the insect order Lepidoptera (moths and butterflies) bind the naturally insecticidal Cry toxins produced by the bacterium *Bacillus thuringiensis*. Phylogenetic analysis of amino acid sequences of seven lepidopteran APN classes provided strong support for the hypothesis that lepidopteran APN2 class arose by gene duplication prior to the most recent common ancestor of Lepidoptera and Diptera. The Cry toxin-binding region (BR) of lepidopteran and dipteran APNs was subject to stronger purifying selection within APN classes than was the remainder of the molecule, reflecting conservation of catalytic site and adjoining residues within the BR. Of lepidopteran APN classes, APN2, APN6, and APN8 showed the strongest evidence of functional specialization, both in expression patterns and in the occurrence of conserved derived amino acid residues. The latter three APN classes also shared a convergently evolved conserved residue close to the catalytic site. APN8 showed a particularly strong tendency towards class-specific conserved residues, including one of the catalytic site residues in the BR and ten others in close vicinity to the catalytic site residues. The occurrence of class-specific sequences along with the conservation of enzymatic function is consistent with the hypothesis that the presence of Cry toxins in the environment has been a factor shaping the evolution of this multi-gene family.

Keywords

aminopeptidase N; convergent/parallel evolution; Diptera; Lepidoptera; multi-gene family

1. Introduction

Gene duplication, followed by functional divergence of duplicated gene copies, is believed to be an important mechanism in the origin of new adaptive phenotypes (Hughes 1994). In the genomes of multi-cellular eukaryotes, many genes belong to multi-gene families that have originated as a result of numerous gene duplication events over evolutionary history (Friedman and Hughes 2001). Phylogenetic methods can be used to reconstruct the history

© 2014 Elsevier Inc. All rights reserved.

Address correspondence to: Austin L. Hughes, Ph.D. Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia SC 29208 USA, Tel: 1-803-777-9186, Fax: 1-803-777-4002, austin@biol.sc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of gene duplication within multi-gene families; and by comparing the phylogenies of genes with organismal phylogenies, it is possible to determine the timing of gene duplication relative to major cladogenetic events, thereby shedding light on the evolution of novel adaptive strategies.

Aminopeptidase N (APN) designates a class of zinc metalloproteinases that preferentially cleave single uncharged amino acids from the N-terminus of polypeptides, which are encoded in insects by the members of a multi-gene family (Nakanishi et al. 2002; Albiston et al. 2004; Piggott and Ellar 2007; Crava et al. 2010). APNs expressed in the midgut of members of the insect order Lepidoptera (moths and butterflies) have been intensely investigated because they bind the Cry toxins produced by the bacterium *Bacillus thuringiensis* (Bt). These toxins are harmless to mammals and thus have potential as natural insecticides (Knight et al. 1995; Bravo et al. 2007). There do not appear to be any published estimates of the time over which *B. thuringiensis* and the insects it infects have co-evolved. However, the extensive diversity of Bt Cry proteins and their toxicity to insects of several different orders suggest a long co-evolutionary history, perhaps hundreds of millions of years (Schnepf et al. 1998; de Maagd et al. 2001). APNs are encoded by the members of a multi-gene family in Lepidoptera (Crava et al. 2010); thus, it is possible that selection arising from Cry toxins has been a factor in the evolution of the insect APN multi-gene family.

The individual susceptibilities of different APN family members to Cry toxins have been investigated experimentally in certain lepidopteran species. For example, in the silkworm *Bombyx mori*, protein fragments corresponding to the toxin-binding regions of four different APN family members were found to bind the Bt toxins Cry1Aa and Cry1ab *in vitro* (Nakashini et al. 2002). However, in the brush border membrane vesicle of *B. mori*, only one intact silkworm APN (APN1) showed detectable binding of the same Bt toxins (Nakanishi et al. 2002). This difference might be attributed to differences in expression, toxin-binding, or both among *B. mori* APNs (Nakashini et al. 2002). Consistent with an important role for APN1 in the toxicity of Bt Cry, resistance to Cry toxins is associated in several lepidopteran species with lack of expression of the APN1 ortholog (Herrero et al. 2005; Zhang et al. 2009).

A number of authors have published phylograms of lepidopteran APNs (Nakanishi et al. 2002; Rajagopal et al. 2003; Angelucci et al. 2008; Crava et al. 2010). Initially, these analyses were somewhat inconsistent in the identification and nomenclature of APN paralogs in Lepidoptera, but much of this confusion was resolved by the thorough analysis of Crava et al. (2010), which identified eight classes of lepidopteran APNs (APN1-8). However, published analyses have not used model-based phylogenetic methods, instead relying on pairwise alignment similarity; and therefore evolutionary relationships among the APN classes remain unclear. In addition, published analyses have often not included APN sequences from non-lepidopteran insect orders which might be used to date APN gene duplications relative to the origin of the Lepidoptera.

Here I conduct a maximum likelihood phylogenetic analysis of APN sequences from Lepidoptera and from Diptera (the flies), whose most recent common ancestor (MRCA)

with Lepidoptera lived in the Permian Period 250-300 million years ago (Wiegmann et al. 2009). In order to examine patterns of functional differentiation among paralogs, I analyze data on gene expression in *B. mori* (Xia et al. 2007), examine patterns of sequence conservation, and reconstruct amino acid sequence changes on the branches leading to major clades of APNs. The goal of reconstructing ancestral amino acids was to identify amino acid residues that are both derived and conserved in each class of lepidopteran APNs. Derived residues (i.e., those that originated in the MRCA of a clade of sequences) that are conserved within that clade are candidates for functional specialization of clade members at the amino acid sequence level (Hughes 2012a, 2012b, 2013).

Gene duplication, followed by functional divergence of duplicated gene copies, is believed to be an important mechanism in the origin of new adaptive phenotypes (Nei M 1969; Hughes 1994). In the genomes of multi-cellular eukaryotes, many genes belong to multi-gene families that have originated as a result of numerous gene duplication events over evolutionary history (Friedman and Hughes 2001). Phylogenetic methods can be used to reconstruct the history of gene duplication within multi-gene families; and by comparing the phylogenies of genes with organismal phylogenies, it is possible to determine the timing of gene duplication relative to major cladogenetic events, thereby shedding light on the evolution of novel adaptive strategies (e.g., Friedman and Hughes 2002; Roelofs and Rooney 2003; Rewitz et al. 2007; Hughes 2012a, 2012b, 2013). Here I apply these methods to gain insight into the functional diversification of lepidopteran APNs.

2. Methods

2.1. Phylogenetic Analysis

The phylogenetic analysis was based on 81 selected APN-family amino acid sequences from Lepidoptera and Diptera. Lepidopteran species included representatives of seven families belonging to four superfamilies; the species used and their classification are listed in Table 1. Diptera were represented by two species of mosquitoes (family Culicidae) and two *Drosophila* species (Table 1). All sequences were downloaded from the NCBI database except for one (BGIBMGA008061), which was obtained from the Silkworm Genome Database (<http://silkworm.genomics.org.cn/silkdb/>). Biochemical assays for APN activity have been applied to certain sequences, particularly from Lepidoptera (Denolf et al. 1997; Rajagopal et al. 2003; Yang et al. 2010), but the majority of sequences have been assigned to the APN family by sequence homology (e.g., Crava et al. 2010).

The Cry toxin-binding domain was identified by homology following Nakanishi et al. (2002). Only sequences including the conserved DEP amino acid sequence motif at the C-terminus of the Cry toxin-binding domain were used in the analysis. The sequences used represented all of the Lepidopteran APN classes identified by Crava et al. (2010) except APN7, which lacks the DEP motif. Since no crystal structure of an insect APN is available, potentially functionally important motifs were identified by aligning insect APNs with human APN (gi157266300), for which a crystal structure is available (Wong et al. 2012). The human APN had a 33.6% mean amino acid sequence identity with the insect APN sequences used in this study.

Amino acid sequences were aligned by the CLUSTAL algorithm in MEGA 5.2 (Tamura et al. 2011); see Supplementary Figure S1. In evolutionary analyses of a set of sequences, any site at which the alignment postulated a gap in any sequence was excluded from the analyses. The maximum likelihood (ML) analysis was based on the WAG+G+I+F model, which was chosen using the Bayes Information Criterion in MEGA 5.2. The reliability of the clustering patterns in the ML tree was tested by bootstrapping; 1000 bootstrap pseudo-samples were used.

The number of synonymous substitutions per synonymous site (d_S) and the number of synonymous substitutions per non-synonymous site (d_N) were estimated by the modified Nei-Gojobori method (Nei and Kumar 2000). This method takes into account transitional bias, which influences estimates of d_S and d_N , particularly at two-fold degenerate sites (Li 1993; Nei and Kumar 2000; Vipin Kumar et al. 2012). We estimated the transition:transversion ratio from the sequences to be compared by the MCL method in MEGA.

Ancestral amino acid sequences (most probable ancestors) were reconstructed by ML in MEGA 5.2. For a given clade in the phylogeny, *conserved derived amino acid residues* were defined as those which were reconstructed to have arisen as amino acid replacements in the branch ancestral to that clade and which were 100% conserved in all members of that clade used in the analysis. Conserved derived amino acid residues are candidates for playing a role in clade-specific functions. Of course not all such replacements will be functionally significant, since some may result from selectively neutral substitutions that are conserved by chance. Obviously, the number of residues that are conserved within a given clade will be in part a function of the number of clade members (i.e., the number of evolutionary lineages) available for analysis, since more variants at functionally unimportant sites are likely to be seen in a larger sample of lineages than in a smaller sample. For this reason, the percentage of conserved sites which were derived was used as a measure of the extent of amino acid sequence specialization of each APN clade, since this percentage is expected to be independent of the number of lineages available for analysis.

2.2. Gene Expression Data

Normalized microarray expression scores of *Bombyx mori* (silkworm) probes corresponding to the *APN1*, *APN2*, *APN3*, *APN4*, *APN6*, and *APN8* genes were downloaded from the NCBI GEO database (accession GSE17571). The data were taken from a study (Xia et al. 2007) that used a custom genome-wide microarray with 22,987 70-mer oligonucleotide probes covering known and predicted *B. mori* genes. For each biological replicate, RNA extracted from 100 silkworms was pooled (Xia et al. 2007). The raw intensity data were normalized by a linear normalization method using four confirmed housekeeping genes as a standard (Xia et al. 2007). For purposes of the present analysis, when the original data set included results for more than one probe from a given gene, scores were averaged across probes. The available data were obtained from replicates of both males and females. Since no sex differences in APN expression were observed, replicates of both sexes were treated as independent biological replicates for purposes of the present analysis. The statistical analyses reported here were conducted in Minitab version 15.0 (<http://www.minitab.com>); all reported P-values are based on two-tailed tests.

3. Results

3.1. Phylogenetic Analysis

A phylogenetic analysis of APNs included clusters corresponding to the lepidopteran APN classes APN1-6 and APN8 (Crava et al. 2010). Each of these clusters received 100% bootstrap support and included only sequences from Lepidoptera not Diptera (Figure 1). Except for the APN8 cluster, each of these clusters included sequences from all four of the lepidopteran superfamilies used in the analyses (Figure 1 and Table 1). The APN8 cluster, by contrast, included only three sequences, representing the superfamilies Pyraloidea and Bombycoidea (Figure 1 and Table 1).

The phylogenetic tree revealed two major clusters including sequences of both Lepidoptera and Diptera: (1) a cluster including APN1, APN3, APN4, APN6, and APN8 of Lepidoptera along with three APN sequences from the dipteran family Culicidae (mosquitoes); and (2) a cluster including APN2 of Lepidoptera along with the sequences from the dipteran families Culicidae and Drosophilidae (Figure 1). The branch separating these two major clusters received 100% bootstrap support (Figure 1). Because each cluster included sequences from both Lepidoptera and Diptera, this topology supported the hypothesis that the gene duplication giving rise to these two clusters took place prior to the MRCA of Lepidoptera and Diptera. Thus, lepidopteran APN2 diverged from the common ancestor of the other lepidopteran APNs analyzed prior to the MRCA of Lepidoptera and Diptera.

The phylogenetic tree supported a sister relationship between APN1 and APN3; the branch supporting this pattern received 100% bootstrap support (Figure 1). Since both APN1 and APN3 clusters included representatives of all four lepidopteran superfamilies from which sequences were available (Table 1 and Figure 1), this topology supported the hypothesis that the gene duplication giving rise to APN1 and APN3 occurred after the origin of Lepidoptera but before the MRCA of these superfamilies. In the phylogenetic tree, a cluster of three mosquito sequences branches next to APN1 and APN3; however, this pattern received only moderate (61%) bootstrap support (Figure 1). If this topology is correct, it would indicate that the ancestor of APN1 and APN3 arose by gene duplication prior to the MRCA of Lepidoptera and Diptera.

The phylogenetic tree supported (99% bootstrap support) a sister-group relationship between APN5 and APN6 (Figure 1). APN4, in turn, clustered as a sister-group to APN5 and APN6, with 100% bootstrap support (Figure 1). APN8 branched off next, but the clustering of APN8 with APN4, APN5, and APN6 received only 68% bootstrap support.

3.2. Gene Expression

Among APNs for which expression data were available, significant differences in expression among genes were seen in the midgut, Malpighian tubule, and head (Figure 2). In all three cases, *APN1* showed the highest expression levels (Figure 2). Tissues differed, however, with respect to the relative expression levels of the other genes. For example, *APN8* showed very low expression in the midgut but much higher expression in the Malpighian tubule (Figure 2). On the other hand, *APN2* showed relatively low expression

levels in both the midgut and the Malpighian tubule but a high expression level in the head (Figure 2).

3.3. Nucleotide Substitution

The number of synonymous substitutions per synonymous site (d_S) and the number of synonymous substitutions per non-synonymous site (d_N) were estimated separately in the Cry toxin binding region (BR) and in the remainder of the gene for orthologous pairs of APN genes from (1) two species of Lepidoptera, *Bombyx mori* and *Ostrinia nubilalis* (Table 2) and (2) two species of Diptera, *Aedes aegypti* and *Anopheles gambiae* (Table 3). In all cases, d_S was much higher than d_N (data not shown), consistent with the prevalence of purifying selection; but because d_S values were near saturation (greater than 1.0 substitutions per site), d_S values are not reported here.

In overall means of all orthologous comparisons between the two Lepidoptera, mean d_N in the BR was significantly lower than that in the remainder of the gene (Table 2). Likewise in the Diptera, mean d_N in the BR was significantly lower than that in the remainder of the gene (Table 3). Thus in both orders, purifying selection was stronger on the BR than on the remainder of the APN protein, indicating that the amino acid sequence of the BR is subject to greater functional constraint than is that of the remainder. However, there was substantial variation among comparisons with respect to the degree of difference between d_N in the remainder and d_N in the BR. In APN6 and APN8 of Lepidoptera, d_N in the remainder was over twice as high as d_N in the BR (Table 2). In no other comparison in either order was the ratio of d_N in the remainder to d_N in the BR as high as 2:1 (Tables 2 and 3).

3.4. Conserved Derived Amino Acid Residues

ML reconstruction of ancestral amino acid sequences was applied to the 577 sites used in the phylogenetic analysis (Figure 1). The seven classes of lepidopteran APNs differed with respect to the proportion of residues conserved in all clade members that were derived; i.e., that arose by mutation in the MRCA of the individual clade (Figure 3A; Supplementary Figure S1). APN1 showed the lowest percentage of conserved residues that were derived, with only 9 of 241 conserved residues (3.7%) being derived (Figure 3A). By contrast, APN8 showed the highest percentage of conserved residues that were derived (26.4%; Figure 3A). The next highest values were seen in APN6 (17.9%) and APN2 (15.9%; Figure 3A). The second lowest value (8.1%) was seen in APN5 (Figure 3A). Thus, conserved amino acid residues in APN1 showed a generalized character, with relatively few of them being specific to the APN1 clade. By contrast, APN8, APN6, and APN2 showed evidence of a high degree of specialization in amino acid sequence, with relatively large proportions of conserved residues being clade-specific.

For the seven lepidopteran APN classes, the number of conserved derived amino acid residues in the Cry toxin-binding region (BR) was plotted against the number of conserved derived amino acid residues in the remainder of the protein (Figure 3B). There was a significant positive correlation between the two values ($r = 0.855$; $P = 0.014$; Figure 3B). APN1 stood apart from the other classes in having no conserved derived residues in the BR and only 9 in the remainder (Figure 3B). APN8 and, to a lesser extent, APN6 were

characterized by high numbers of conserved derived residues both within and outside the BR (Figure 3B). The remaining classes have intermediate numbers of conserved derived residues in the BR and in the remainder of the protein (Figure 3B).

The amino acid residues involved in the APN catalytic site of human APN include the amino acid motif GAMEN (residues 352-356 of the human sequence; Wong et al. 2012). This motif was completely conserved in all lepidopteran APNs analyzed except APN5 and APN6, which have GATEN. Reconstruction of ancestral amino acids indicated that the amino acid replacement M → T in the central residue of this motif occurred in the branch leading to the common ancestor of APN5 and APN6.

Additional residues in the catalytic site of human APN include two that are in the region of the molecule homologous to the BR (residues Q211 and Q213 of human APN; Wong et al. 2012). These residues are located at the C-terminal end of a β -strand and in the hydrogen-bonded turn that follows it (Wong et al. 2012; Table 4). This region included conserved derived amino acid residues in several lepidopteran APN classes (Table 4). APN8 was unique in having five conserved derived residues in this region, including one of the two catalytic site residues (Table 4). APN2 and APN6 each included two conserved derived residues in this region, while APN3 and APN5 each included one conserved derived residue in this region (Table 4). By contrast APN1 and APN4 included no conserved derived residues in this region (Table 4). No other lepidopteran APN class besides APN8 included a conserved derived residue at one of the catalytic site positions (Table 4).

Catalytic site residues in human APN also include several residues in a helical region where the zinc ion-binding residues are located (Table 4). The corresponding region of lepidopteran APNs included class-specific conserved derived residues in each of the seven classes (Table 4). There were six conserved derived residues in this region of APN8 and three each in APN5, APN6, and APN2 (Table 4). Of particular interest was a threonine replacement located at a site (corresponding to L407 in human APN) which is located close to the zinc-ion binding residues (E411 in human APN). This residue was predicted to have arisen by replacement of an ancestral leucine in the branches leading to APN8 and APN8 and by replacement of an ancestral valine on the branch leading to APN6. Thus, the conserved derived threonine residue at this position represents a case of parallel/convergent evolution at the amino acid sequence level.

4. Discussion

A phylogenetic analysis of amino acid sequences was used to test hypotheses regarding the evolutionary relationships and time of origin of seven classes of aminopeptidase N (APN) recognized in the order Lepidoptera. The phylogenetic tree provided strong support for the hypothesis that lepidopteran APN2 class arose by gene duplication prior to the MRCA of Lepidoptera and Diptera, which occurred in the Permian Period 250-300 million years ago (Figure 4; Wiegmann et al. 2009). Weaker support was provided for the hypothesis that the gene duplication separating the common ancestor of Lepidopteran APN1 and APN3 from the common ancestor of Lepidopteran APN4, APN5, APN6, and APN8 also occurred prior to the MRCA of Lepidoptera and Diptera.

Lepidopteran APN1 and APN3 were sister groups in the phylogenetic analysis. Since no dipteran orthologs of either APN1 or APN3 were discovered, the duplication separating APN1 from APN3 must have occurred after the MRCA of Lepidoptera. Both APN1 and APN3 clusters included members of both the superfamily Yponomeutoidea and the superfamilies Bombycoidea, Pyraloidea, and Noctuoidea (Figure 1 and Table 1). Therefore, the gene duplication leading to separate *APN1* and *APN3* genes must have occurred before the MRCA of these superfamilies, which is believed to have occurred about 80 million years ago (Figure 4; Grimaldi and Engel 2005). Lepidopteran APN5 and APN6 likewise were sister groups in the phylogenetic analysis. Since an APN5 sequence was available from Yponomeutoidea, the phylogeny supported the hypothesis that *APN5* and *APN6* genes duplicated before the MRCA of the four subfamilies. Since the phylogenetic tree indicated that the *APN4* and the *APN8* genes diverged before the duplication of the *APN5* and *APN6* genes, these genes also must have originated by gene duplication prior to the MRCA of the four superfamilies from which sequences were available.

There was evidence that functional divergence among paralogous APN genes in Lepidoptera was multi-dimensional, including both differences in gene expression and in amino acid sequence. In microarray data from *B. mori*, *APN1* showed relatively high expression levels across tissues, whereas other *APN* genes showed more tissue-specific expression patterns, especially *APN8* and *APN2* (Figure 2). Results from qRT-PCR experiments in *O. nubilabilis* by Crava et al. (2010), like the *B. mori* microarray data analyzed here (Figure 2B), showed a high level of *APN8* expression in the Malpighian tubule. Thus at least some patterns of specialized gene expression of lepidopteran APNs may be ancient, predating the MRCA of the four lepidopteran superfamilies analyzed here.

In comparisons of orthologous APNs within both Lepidoptera and Diptera, there was evidence that functional constraint on the amino acid sequence were greater in the Cry toxin binding region (BR) than in the remainder of the protein. This constraint was particularly strong in *APN6* and *APN8* of Lepidoptera. The evidence of constraint on the BR suggests that amino acid residues in this region are functionally important. The functional role of these residues may relate to the fact that, as inferred from homology with the structure of human APN (Wong et al. 2012), certain residues located in the BR are predicted to contribute to the catalytic site of the enzyme.

Sequence divergence among the lepidopteran APN classes involved numerous class-specific conserved derived amino acid replacements. Although derived residues that are conserved among a clade of sequences often include amino acid replacements that are selectively neutral, were fixed by genetic drift, and are subsequently conserved by chance, conservation of a residue is consistent with purifying selection and thus with functional importance. Thus, conserved derived amino acid replacements provide candidate residues for class-specific functions. Potential functional importance of some conserved derived residues in the case of lepidopteran APNs was provided by the evidence of parallel/convergent evolution of a conserved threonine replacement near the catalytic site of the enzyme which occurred in the ancestry of *APN6*, *APN8*, and *APN2* (Table 4). Independent occurrence of the amino acid replacement in three different lineages coupled with within-class conservation is strongly suggestive of functional importance.

Analysis of class-specific conserved derived residues indicated that specialization in amino acid sequence has not occurred uniformly across lepidopteran APN classes. The highest proportions of conserved derived residues were seen in APN6, APN8, and APN2. APN8 and APN6 showed particularly high numbers of conserved derived residues in the BR, and in the case of APN8 these residues included one of the catalytic site residues and several others in the near vicinity of the catalytic site residues (Table 4). APN8 also showed numerous other conserved derived residues near the catalytic site, where APN6 and APN2 also showed relatively high numbers of conserved derived residues (Table 4). Interestingly, the APN classes with the strongest evidence of amino acid sequence specialization included the APN classes with the strongest within-class conservation of the BR (APN6 and APN8) and those with most evidence of tissue-specificity in gene expression in *B. mori* (APN8 and APN2). Thus, in the case of APN8 and APN2, specialization in amino acid sequence was associated with specialization in expression pattern as well.

APN1, the most broadly expressed of the lepidopteran APNs, showed the least evidence of sequence specialization, having no conserved derived residues in the BR and only nine in the rest of the sequence. Thus APN1 appears to represent the most functionally generalized class of lepidopteran APN. APN2, APN3, APN4, and APN5 showed relatively little evidence of sequence specialization in the BR, while showing greater evidence of sequence specialization outside the BR. This observation is of interest in the light of evidence that Cry toxins bound protein fragments corresponding to the BR of *B. mori* APN1, APN2, APN3, and APN4 members (Nakanishi et al. 2002).

On the other hand, the observation that the same toxins only detectably bound intact APN1 *in vivo* implies that the factors other than the BR sequence itself account for Cry toxin binding of the intact molecule (Nakanishi et al. 2002). Amino acid sequence changes outside the BR might contribute to reduced susceptibility to Cry toxins, perhaps by affecting the three-dimensional structure of the molecule in such a way as to render the BR less accessible to Cry toxins. Information on the three-dimensional structures of lepidopteran APN classes and functional studies will be needed to test this hypothesis. By identifying class-specific conserved amino acid replacements, the present analyses suggest candidate residues for class-specific functions, including class-specific Cry toxin susceptibility. The present results are thus consistent with the hypothesis that the APN family of Lepidoptera have evolved subject to selective pressures originating both from their enzymatic function and from the environmental presence of Cry toxins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Albiston AL, Ye S, Chai SY. Membrane bound members of the M1 family more than aminopeptidases. *Protein Pept Lett*. 2004; 11:4921–500.
- Angelucci C, Barrett-Wilt GA, Hunt DF, Akhurst RJ, East PD, Gordon KH, Campbell PM. Diversity of aminopeptidases, derived from four lepidopteran gene duplications, and polycalins expressed in

- the midgut of *Helicoverpa armigera*: identification of proteins binding the δ -endotoxin, Cry1Ac of *Bacillus thuringiensis*. *Insect Biochem Mol Biol*. 2008; 38:685–696. [PubMed: 18549954]
- Bravo A, Gill SS, Soberon M. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon*. 2007; 49:423–435. [PubMed: 17198720]
- Crava CM, Bel Y, Lee SP, Manachini B, Heckel DG, Escriche B. Study of the aminopeptidase N gene family in the lepidopterans *Ostrinia nubilalis* (Hübner) and *Bombyx mori* (L.): sequences, mapping and expression. *Insect Biochem Mol Biol*. 2010; 40:506–515. [PubMed: 20420910]
- de Maagd R, Bravo A, Crickmore N. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. *Trends Genet*. 2001; 17:193–199. [PubMed: 11275324]
- Friedman R, Hughes AL. Gene duplication and the structure of eukaryotic genomes. *Genome Research*. 2001; 11:373–381. [PubMed: 11230161]
- Friedman R, Hughes AL. Molecular evolution of the NF- κ B signaling system. *Immunogenetics*. 2002; 53:964–974. [PubMed: 11862396]
- Grimaldi, D.; Engel, MS. *Evolution of the Insects*. Cambridge: Cambridge University Press; 2005.
- Herrero S, Gechev V, Bakker PL, Moar WJ, de Maagd RA. *Bacillus thuringiensis* Cry1Ca-resistant *Spodoptera exigua* lacks expression of one of four aminopeptidase N genes. *BMC Genomics*. 2005; 6:96. [PubMed: 15978131]
- Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B*. 1994; 256:119–124.
- Hughes AL. Evolution of the β GRP/GNBP/ β -1,3-glucanase family of insects. *Immunogenetics*. 2012a; 64:549–558. [PubMed: 22410636]
- Hughes AL. Evolution of the heme peroxidases of Culicidae (Diptera). *Psyche*. 2012b; 2012:146387.
- Hughes AL. Evolution of the salivary apyrases of blood-feeding arthropods. *Gene*. 2013; 527:123–130. [PubMed: 23791653]
- Knight PJ, Crickmore N, Ellar DJ. The receptor for *Bacillus thuringiensis* Cry1A(c) delta-endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase. *N Mol Microbiol*. 1994; 11:429–436.
- Li WH. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J Mol Evol*. 1993; 36:96–99. [PubMed: 8433381]
- Nakanishi K, Yaoi K, Nagino Y, Hara H, Kitami M, Atsumi S, Miura N, Sato R. Aminopeptidase N isoforms from the midgut of *Bombyx mori* and *Plutella xylostella* – their classification and the factors that determine their binding specificity to *Bacillus thuringiensis* Cry1A toxin. *FEBS Lett*. 2002; 519:215–220. [PubMed: 12023048]
- Nei M. Gene duplication and nucleotide substitution in evolution. *Nature*. 1969; 221:40–42. [PubMed: 5782607]
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986; 3:418–426. [PubMed: 3444411]
- Nei, M.; Kumar, S. *Molecular evolution and phylogenetics*. Oxford University Press; New York: 2000.
- Piggott CR, Ellar DJ. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. *Microbiol Mol Biol Rev*. 2007; 71:255–281. [PubMed: 17554045]
- Rajagopal R, Agrawal N, Selvandandian A, Sivakumar S, Ahmad S, Bhatnagar RK. Recombinantly expressed isoenzymatic aminopeptidases from *Helicoverpa armigera* (American cotton bollworm) midgut display differential interaction with closely related *Bacillus thuringiensis* insecticidal proteins. *Biochem J*. 2003; 370:971–978. [PubMed: 12441000]
- Rewitz KF, O'Connor MB, Gilbert LI. Molecular evolution of the insect Halloween family of cytochrome p450s: phylogeny, gene organization and functional conservation. *Insect Biochem Mol Biol*. 2007; 37:741–753. [PubMed: 17628274]
- Roelofs WL, Rooney AP. Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *Proc Natl Acad Sci USA*. 2003; 100:9179–9184. [PubMed: 12876197]
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011; 28:2731–2739. [PubMed: 21546353]

- Vipan Kumar S, Apurba D, Singh A. Comparative analysis of nonsynonymous and synonymous substitution of capsid proteins of human herpes virus. *J Proteomics Bioinform*. 2012; 5:172–176.
- Wiegmann BM, Trautwein MD, Kim JW, Cassel BK, Bertone MA, Winterton SL, Yeates DK. Single-copy nuclear genes resolve the phylogeny of holometabolous insects. *BMC Biol*. 2009; 7:34. 2009. [PubMed: 19552814]
- Wong AH, Zhou D, Rini JM. The X-ray crystal structure of human aminopeptidase N reveals a novel dimmer and the basis for peptide processing. *J Biol Chem*. 2012; 287:36804–36813. [PubMed: 22932899]
- Xia Q, Cheng D, Duan J, Wang G, Cheng T, Zha X, Liu C, Zhao P, Dai F, Zhang Z, He N, Zhang L, Xiang Z. Microarray-based gene expression profiles in multiple tissues of the domesticated silkworm, *Bombyx mori*. *Genome Biol*. 2007; 8:R:162. 2007. [PubMed: 17683582]
- Zhang X, Wheeler MM, Oi FM, Scharf ME. Mutation of an aminopeptidase N gene is associated with *Helicoverpa armigera* resistance to *Bacillus thuringiensis* Cry1Ac toxin. *Insect Biochem Mol Biol*. 2009; 39:421–429. [PubMed: 19376227]

Highlights

- The aminopeptidase N (APN) genes of Lepidoptera underwent ancient duplications.
- APN1 was the most generalized APN class.
- Class-specific conserved derived residues suggest sequence specialization.
- APN2, APN6, and APN8 were most specialized in sequence.
- APN2, APN6, and APN8, shared a convergently evolved residue near the active site.

Figure 1. Maximum likelihood tree of APNs from Lepidoptera and Diptera based on WaG+G+I+F model at 577 aligned amino acid positions. Numbers on the branches represent the percentage of bootstrap samples supporting the branch; only values $\geq 50\%$ are shown.

A)

B)

C)

Figure 2.

Mean expression score of *Bombyx mori* transcripts corresponding to APN genes in (A) mid-gut (one-way ANOVA; $F_{5, 18} = 96.35$; $P < 0.001$); (B) Malpighian tubule (one-way ANOVA; $F_{5, 42} = 12.55$; $P < 0.001$); (C) head (one-way ANOVA; $F_{5, 18} = 26.33$; $P < 0.001$).

A)

B)

Figure 3.

(A) Percent of conserved residues that are derived in Lepidopteran APN classes. A residue is counted as conserved if conserved in all members of the class analyzed in the phylogeny (Figure 1); numbers above each bar represent the number of conserved residues in each class. The difference in proportions among classes was highly significant ($\chi^2 = 74.7$; 6 d.f.; $P < 0.001$). (B) Number of conserved derived residues in the Cry toxin-binding region (BR) vs. the remainder of the protein ($r = 0.855$; $P = 0.014$).

Figure 4.

Schematic phylogeny of the lepidopteran superfamilies included in the phylogenetic analyses (after Grimaldi and Engel 2005), showing the reconstructed time of the gene duplications (A) giving rise to *APN2*; and (B) the duplications of *APN1* and *APN3* and the duplications of *APN8*, *APN4*, *APN5*, and *APN6*.

Table 1

Insect species from which aminopeptidase N sequences are analyzed.

Order	Superfamily	Family	Species ^I
Lepidoptera	Yponomeutoidea	Plutellidae	Diamondback moth <i>Plutella xylostella</i>
		Crambidae	Striped rice borer <i>Chilo suppressalis</i>
	Pyraloidea		Rice leaf-folder <i>Cnaphalocrocis medinalis</i>
			Sugar cane borer <i>Diatraea saccharalis</i>
			European corn borer <i>Ostrinia nubilalis</i>
		Nymphalidae	Monarch butterfly <i>Danaus plexippus</i>
		Pyralidae	Indian meal moth <i>Plodia interpunctella</i>
	Bombycoidea	Bombycidae	Silkworm moth <i>Bombyx mori</i>
	Noctuoidea	Lymantriidae	Gypsy moth <i>Lymantria dispar</i>
		Noctuidae	Cotton bollworm <i>Helicoverpa armigera</i>
			Native budworm <i>Helicoverpa punctigera</i>
			Beet armyworm <i>Spodoptera exigua</i>
			Cabbage looper <i>Trichoplusia ni</i>
Diptera	Culicoidea	Culicidae	<i>Aedes aegypti</i>
			<i>Anopheles gambiae</i>
	Ephydroidea	Drosophilidae	<i>Drosophila melanogaster</i>
			<i>Drosophila grimshawi</i>

^I Common names are included for Lepidoptera.

Table 2

Number of nonsynonymous substitutions per nonsynonymous site ($d_N \pm \text{S.E.}$) in Cry toxin binding region (BR) and in the remainder of the gene in comparisons between *Bombyx mori* and *Ostrinia nubilalis* APN orthologs.

Gene	BR	Remainder
<i>APN1</i>	0.2187 ± 0.0436	0.2949 ± 0.0141
<i>APN3</i> ¹	0.2828 ± 0.0536	0.3020 ± 0.0161
<i>APN4</i>	0.2765 ± 0.0510	0.3415 ± 0.0159
<i>APN5</i>	0.2102 ± 0.0414	0.2716 ± 0.0137
<i>APN6</i>	0.1117 ± 0.0316	$0.2946 \pm 0.0148^{***}$
<i>APN8</i>	0.1405 ± 0.0338	$0.3232 \pm 0.0156^{***}$
<i>APN2</i>	0.1887 ± 0.0406	0.2681 ± 0.0136
Mean	0.1966 ± 0.0234	$0.2877 \pm 0.0097^{\dagger\dagger}$

¹ Mean of comparisons of *B. mori* *APN3* with *O. nubilalis* *APN3a* (FJ896130) and with *APN3b* (FJ492806).

*** Z-tests of the hypothesis that d_N in the remainder equals that in BR: $P < 0.001$.

^{††} Paired t-test of the hypothesis that mean d_N in the remainder equals that in BR: $P < 0.01$.

Table 3

Number of nonsynonymous substitutions per nonsynonymous site ($d_N \pm \text{S.E.}$) in Cry toxin binding region (BR) and in the remainder of the gene in comparisons between *Aedes aegypti* and *Anopheles gambiae* putative APN orthologs.

Comparison	BR	Remainder
gi157111303 vs. gi247970418	0.2370 ± 0.0458	$0.3565 \pm 0.0165^*$
gi157111299 vs. gi347970406	0.2327 ± 0.0459	$0.4549 \pm 0.0200^{***}$
gi157120775 vs. gi347971145	0.3716 ± 0.0632	$0.5168 \pm 0.0220^*$
gi157118042 vs. gi 347966742	0.2997 ± 0.0544	$0.5224 \pm 0.0228^{***}$
gi157133543 vs. gi158297815	0.2837 ± 0.0544	0.3779 ± 0.0166
Mean	0.2849 ± 0.0253	$0.4457 \pm 0.0343^{\dagger\dagger}$

* Z-tests of the hypothesis that d_N in the remainder equals that in BR: $P < 0.05$.

*** Z-tests of the hypothesis that d_N in the remainder equals that in BR: $P < 0.001$.

$\dagger\dagger$ Paired t-test of the hypothesis that mean d_N in the remainder equals that in BR: $P < 0.01$.

Table 4

Amino acid sequences of *B. mori* APNs (with variants found in sequences of the same class from other Lepidoptera) in two catalytic site regions.

Class	β -strand and hydrogen-bonded (within Cry toxin-binding region) ^{1,2}	Zinc Ion-Binding Helical Region ²
APN1	WMGTT Q F Q PGH A SY	KVAVTVA ME SLA MD FGNLTTCAMGNLWLA ME I I S T
APN3	WMAST Q F Q ATS Y T HL A T	LIAVTL ME SLA MD FGNLTTCAMGNLWLA ME R S F NE AF A S
APN4	YYATT Q F Q PFH E YY I V	YIAVTL ME SLA MD FGNLTTCAMGNLWLA ME SSL MG V
APN5	WYAATH L Q PTN H	YIGTTL ME SLA MD FGNLTTCAMGNLWLA ME S G T
APN6	IYATTH F Q PYN F	YIAVTL ME SLA MD FGNLTTCAMGNLWLA ME S T F H T
APN8	WLGAT Q ME PTH	KVAQIVAV ME SLA MD FGNLTTCAMGNLWLA ME I I I I I I
APN2	NLVT S Q L Q PTF KVIA R S	MEAVTICHEN MD FGNLTTCAMGNLWLA ME Q G S MD F S D F T Q V Y

¹The putative hydrogen-bonded turn region is indicated by *italics*.

²Variant residues occurring in other available members of a given class are shown below the *B. mori* sequence. The catalytic site residues are **bold-faced**. Zinc ion-binding residues are underlined. Arrows indicate class-specific conserved derived residues.