Diversification and independent evolution of Troponin C genes in insects

Raul Herranz, Jesus Mateos¹ and Roberto Marco^{*}

Departamento de Bioquímica UAM e Instituto de Investigaciones Biomédicas “Alberto Sols” UAM-CSIC, Facultad de Medicina de la Universidad Autónoma, C/ Arzobispo Morcillo 4, 28029 Madrid. Spain

^{*} Corresponding author: Phone: +034 91 497 54 09, Fax: +034 91 585 44 01

Email: rmarco@uam.es

¹ Present address: Molecular Biology Program, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA

KEYWORDS: TroponinC independent convergence evolution *Drosophila Anopheles* *Apis* subfunctionalization

RUNNING HEAD: TpnC gene evolution in Insects

ABBREVIATIONS: Troponin (Tpn), Dm (*D. melanogaster*), Dp (*Drosophila pseudoobscura*), Ds (*Drosophila subobscura*), Dv (*Drosophila virilis*), Ag (*Anopheles gambiae*), Am (*Apis mellifera*), IFM (Indirect Flight Muscles), TDT (Tergal Depressor of the Trochanter muscle), LMCA (Late Metazoan Common Ancestor), RT-PCR (Reverse Transcription and Polymerase Chain Reaction), MY (Million Years).

Troponin C gene evolution in insects

Troponin C (TpnC), the calcium-binding subunit of the troponin regulatory complex in the muscle thin filament is encoded by multiple genes in insects. To understand how TpnC genes have evolved, we have characterized gene number and structure in a number of insect species. The TpnC gene complement is five genes in Drosophilidae as previously reported for *D. melanogaster*. Gene structures are almost identical in *D. pseudoobscura*, *subobscura* and *virilis*. Developmental patterns of expression are also conserved in *Drosophila subobscura* and *virilis*. Similar, but not completely equivalent, TpnC gene repertoires have been identified in *Anopheles gambiae* and *Apis mellifera* genomes. Insect TpnC sequences can be divided in three groups, allowing a systematic classification of newly identified genes. The pattern of expression of the *Apis mellifera* genes essentially agree with the pattern in Drosophilidae, providing further functional support to the classification. A model for the evolution of the TpnC genes is proposed including the most likely pathway of insect TpnC diversification. Our results suggest that the rapid increase in number and sequence specialization of the adult Type III isoforms can be correlated with evolution of the holometabolous mode of development and the acquisition of asynchronous indirect flight muscle function in insects. This evolutionarily specialization has probably been achieved independently in different insect orders.

INTRODUCTION

Myogenesis is characterized by the regulated expression in different muscle types of a large spectrum of isoforms, many of essentially unknown function, encoded by structural and regulatory genes. Troponin is a three subunit complex that binds to tropomyosin, mainly through its larger T subunit (for reviews, Filatov et al. 1999; Gordon et al. 2000). TpnC senses increases in the intracellular levels of calcium and induces a conformational change in the tropomyosin-troponin complex (Reinach et al. 1997; Vassilyev et al. 1998) that triggers the contraction event. The most extensive studies on TpnC to date have been performed in mammals where the main functional characteristics of the troponins in the sarcomere thin filament have been established. TpnC shows a characteristic dumbbell structure, two globular domains containing two calcium binding EF-Hands each linked by an α -helix (Grabarek et al. 1992). The carboxy terminal domain has a more structural role, since its highly conserved third and fourth EF-Hands are constitutively occupied by cations. The more variable amino-terminal regulatory EF-Hands show lower affinity for calcium and are cation free under basal conditions (Herzberg and James 1985).

Many proteins share this calmodulin-like structure, including the myosin light chains (da Silva et al. 1995), another component of the muscle contraction apparatus. Since the seminal studies of Kretsinger on EF-hand proteins (Kawasaki et al. 1998; Nakayama and Kretsinger 1993; Nakayama et al. 1992), it has become evident that the known TpnCs belonged to different groups. Kretsinger included the vertebrate TpnCs in the CTER superfamily using the initials of proteins in the family (Calmodulins, vertebrate TpnCs, the Essential and the Regulatory myosin light chain proteins), while invertebrate TpnCs were treated as an independent group, characterized by the loss of cation binding ability in some of their EF-Hands.

Troponin C gene evolution in insects

Dipterans such as the *Drosophilidae* and *Anopheles gambiae*, and also many additional holometabolous insects such as *Apis mellifera* (honey bee), have two different sets of musculatures, one formed in the embryos for the larvae and a second, formed during pupation for the imagoes. Many of the adult muscles are physiologically unique to that stage of development including the asynchronous fibrillar indirect flight muscles (IFM) and the tergal depressor of the trochanter (TDT). The latter, which extends into the second leg, is responsible for the jump at take-off of the fly. Myofibrillar protein isoform diversity can be generated via use of multigene families (eg actin, Lovato et al., 2001; TpnC, Qiu et al., 2003, Herranz et al., 2004), but is more usually obtained by differential splicing of a single gene (Bernstein et al. 1993). Given the great diversity of muscle types and life histories within the insect orders we sought to determine if changes in the number and structure of myofibrillar protein genes might respond to changes in muscle complexity and life history.

The number and distribution of troponin genes in metazoans are highly variable. In mammals, troponin I and troponin T are encoded by three genes each, organized in TnT-TnI pairs (Barton et al. 1999; Huang and Jin 1999), while in insects each is encoded by a single gene (Barbas et al 1993, Benoist et al, 1998). Troponin C is encoded by only two genes in mammals despite the repeated genome duplication events described in the vertebrate evolutionary line (Meyer and Schartl 1999; Ohno 1999). Recently, the complete TpnC gene repertoire in *D. melanogaster* was identified, increasing their number from three (Fyrberg et al. 1994) to five genes (Qiu et al. 2003; Herranz et al. 2004).

In the present article, we have studied the properties of TpnC genes in the insecta, namely in Diptera (several drosophilids and *Anopheles*) and in Hymenoptera (*Apis*), covering an evolutionary range of more than 300 MY. We address three main questions: 1) how many genes encode for TpnC in insects, 2) how we can comprehensively classify them

Troponin C gene evolution in insects

and 3) how these genes may have evolved to reach their specific functional properties. The expression patterns in Drosophilidae and *Apis* are also reported and discussed in terms of the classification. The analysis suggests a high complexity and plasticity in the evolution and functional specialization of this gene family.

MATERIALS and METHODS

Insect stocks— *D. melanogaster* Oregon R strain was used as wild-type control set. *D. subobscura* was obtained from Dr. Rosa de Frutos (Valencia, Spain) and *D. virilis* from Dr. Manuel Calleja (Madrid, Spain). These species diverged from *D. melanogaster* around 12 and 50 millions of years ago (Gaunt and Miles 2002; Powell 1997). Samples from *Apis mellifera* were obtained in the summer 2003 from Industrias Alonso located in El Vellon (Madrid). We have included also the genomic data obtained from *D. pseudoobscura* (diverged from *D. subobscura* 5-10 MY ago) and *A. mellifera* (diverged from dipterans more than 300 MY ago) from the Human Genome Sequencing Project at Baylor College of Medicine Blast server. Finally we included the mosquito *Anopheles gambiae* genome data (diverged from Drosophilidae around 250 MY ago, Holt et al. 2002) and the recently published information on the TpnCs of *Lethocerus* (Qiu et al. 2003), the giant water bug, an hemipterous insect. See also supplementary material Figure A.

TpnC genes sequencing and Expression Pattern analysis— Nucleid acid extraction, reverse transcriptions, cloning and sequencing were performed with *D. subobscura*, *D. virilis* and *A. mellifera* as described previously for *D. melanogaster* (Herranz et al. 2004). Specific probes, listed below, were designed for each gene amplification, based in known sequences from *D. melanogaster* and *A. mellifera* genomes. Except for DvTpnCIb gene sense primer (+, 5'-GGC AGT ATC CCC ACC GAA AT-3'), the same probes can be used in all the

Troponin C gene evolution in insects

Drosophila species studied. DmTpnCIa: (+) 5'-TAC ACC CGA GCA GAT CGC CGT-3', (-) 5'-ACT CGC CAG TCA TCA TCT CCA-3'. DmTpnCIb: (+) 5'-GAC GAA GAC CTG ACC CCC GAG C-3', (-) 5'-GGC CTC ATC ATC CTC CTC TA-3'. DmTpnCII: (+) 5'-ATG GAC ATC ATG CGC AAG GCA TTC-3', (-) 5'-ACT CGC CAG TCA TCA TCT CCA-3'. DmTpnCIIIa: (+) 5'-ATG AAT TGA CTA AGG ARC ARA C-3', (-) 5'-CAT AAA TTC ATC AAA RTC NAC-3'. DmTpnCIIIb: (+) 5'-GCT TTC GAC CAG GAC GGT GCA GGA-3', (-) 5'-GAA ATT GAT GCG GAT GGA TCC GGT-3'. AmTpnCI: (+) 5'-TGA CAG GGA GAA AAG TGG TAG-3', (-) 5'-CTG TTA ATT GAT CGT CCA GTT C-3'. AmTpnCIIa: (+) 5'-TTC CGA CTT GAA TGC TTT GAT-3', (-) 5'-ATT TAT TCT CCA GTC ATC ATT TCC-3'. AmTpnCIIb: (+) 5'-GAA GGA AAA AGT GAG GAC GA-3', (-) 5'-AAT AAT AGG AGG AAG AGA GTG TG-3'. AmTpnCIIIa: (+) 5'-GAC AAA AAA GGT AGT ATC GGT-3', (-) 5'-CGG TCA TAA CTT CCA TAA ATT-3'. AmTpnCIIIb: (+) 5'-GAG ATT CAA CTC AGC AAA GAT-3', (-) 5'-GGT GAT AAT TTA GCC TGT CAT-3'.

Bioinformatic Tools and Sequence Accession Numbers— Sequences were obtained and initially compared with NCBI BLAST (Altschul et al. 1997) and sequence information was processed using GeneJockey II® software (Biosoft, Cambridge CB2 1NS, UK). Additional information is provided in the figure legends. Sequences obtained or annotated by us have been submitted to the corresponding section of the Genbank with the following accession numbers: *D. suboscuro* and *D. virilis* AY292251-60, *D. pseudoobscura* BK001456-60 and *A. mellifera* BK001609-12. This information has been prepared also in a Supplementary Material Table A which also includes the accession numbers of the rest of sequences used in this work. Note that the *Anopheles* type III cluster is not yet well annotated in the

Anopheles databases. The correct annotation can be found in the Genbank under accession number BK001613.

RESULTS

Identification of the complete set of TpnC genes in selected Ecdysozoa

To identify the corresponding putative TpnC genes, we started by screening the *D. melanogaster*, *A. gambiae* and *C. elegans* completely sequenced genomes. Using the most evolutionarily conserved TpnC protein sequence for each of the model organisms in BLAST searches (Altschul et al. 1997), we obtained the 15 best score proteins with four EF-Hand motifs in each genome. By representing the degree of identity/similarity to the TpnC used to recover the sequences and to calmodulin, we could easily classify these proteins into Calmodulin-like, TpnC-like and miscellanea (see Supplementary Material, Figure B).

In the three species, we detected one main calmodulin isoform and a second calmodulin-related protein, the androcalmodulin (CG17769, expressed in testis) in the case of *D. melanogaster* (Fyrberg et al. 1994), and the orthologous proteins CE27325 and agCP6219 in *C. elegans* and *A. gambiae*, respectively. Moreover, the TpnC group identified in this analysis confirms that the *C. elegans* genome only contains two TpnC genes (already annotated as TpnC1(pat-10) and TpnC2, (Terami et al. 1999)), while in *A. gambiae* the TpnC candidate gene number increases to six. Five genes form the *D. melanogaster* TpnC repertoire, three earlier described ones (Fyrberg et al. 1994) and two recently identified ones (Qiu et al. 2003; Herranz et al. 2004). Using all properties analyzed there, we grouped the *D. melanogaster* TpnCs in three types: embryonic (I), adult (III) and intermediate (II). When more than one gene falls in one group, we differentiated them with small letters, starting from more widely expressed genes (a) to more specific ones (b). *Apis mellifera* TpnC genes

Troponin C gene evolution in insects

sequences were obtained also from the Baylor College genome server, allowing us to detect five TpnC genes in this species. The troponin gene number comparison among different species is shown in Table I.

To bridge the evolutionary distance between *Drosophila* and *Anopheles*, we cloned the cDNAs encoding the five TpnC genes from *D. subobscura* and *D. virilis* and the orthologous genes of *Apis*, by designing specific probes near the beginning or the end of their coding regions, and obtained sequences that cover the cDNA of the genes (at least 80-90% of the coding regions). For TpnCIb, TpnCII and TpnCIIIb the intronic sequences were sequenced from each of the drosophilid species. All five *Drosophila pseudoobscura* TpnC genes orthologues were found by an isoform-specific BLAST search on the Baylor College genome web server.

Comparative analysis of the TpnC coding sequences

Figure 1 shows the relations between the 32 TpnC sequences obtained from *Apis*, *Anopheles* and drosophilids, already renamed based on their primary sequence homology to their *D. melanogaster* counterparts. This protein-based sequence similarity tree supports the view that there are three gene groups in these insects: 1) a Type I group including drosophilid TpnCIa and Ib, AgTpnCI and AmTpnCI; 2) a Type II group including the drosophilid TpnCII, AgTpnCII, AmTpnCIIa and AmTpnCIIb (less conserved than IIa) and 3) a more heterogeneous Type III group containing the rest of the genes. This group can be subdivided in a IIIa subgroup, relatively well conserved, containing the drosophilid TpnCIIIa, AgTpnCIIIa and AmTpnCIIIa genes and the IIIb subgroup, the less conserved gene set, containing the drosophilid TpnCIIIb, their related *Anopheles* paralogous genes AgTpnCIIIb1, IIIb2 and IIIb3 and the AmTpnCIIIb gene. Type IIIb genes have diverged

Troponin C gene evolution in insects

faster than others and are not phylogenetically grouped. As can be seen in the cDNA based tree included in the Supplementary Material Fig A and C, the putative TpnCIIIb isoform in *Apis* is less closely related to their *Anopheles* and *Drosophila* counterparts than to AmTpnCIIIa. Although the TpnC genes phylogeny cannot be completely deduced from these trees, especially as a consequence of the low bootstrap values near their origins, it is informative enough to classify the genes in three groups. The usefulness of this classification will be seen in the following analysis of their additional functional properties.

Intron indels as footprints of TpnC genes evolution

Yuasa and Takagi (Yuasa et al. 1998; Yuasa and Takagi 2000, 2001) have emphasized the high conservation of relative intron positions during TpnC genes evolution in prochordates and vertebrates. They pointed out that only the third EF-Hand intron position showed variability. Only a single *Drosophila* TpnC gene, TpnCIb (47D), which fitted with their proposal, was included in their work. The study of the conservation of rare genomic events within a gene family, such as intron indels, is useful in analyzing the evolutionary history of genes (Rokas and Holland 2000; Venkatesh et al. 2001). In Figure 2 we show the variability of exon-intron structures of all the sequenced TpnC genes from *A. mellifera*, *A. gambiae* and *Drosophila melanogaster* species. We also include intron identification codes as defined by Nakayama and Kretsinger (1993) (Figure 2 legend, and Supplementary material Table B). Including data from Drosophilidae, *A. gambiae* and *A. mellifera* genes, the TpnC intron repertoire can be also assigned to three classes (figure 2). The more relevant changes in intron positions are:

Troponin C gene evolution in insects

- 1- The perfectly conserved 2.13/1 position in deuterostomes is less conserved in invertebrates. In insects, it is found only in the genes evolved from the Class 1 precursor.
- 2- The 1.01/1 intron is relatively conserved, but it is absent in Type II genes and in Type I genes of the Drosophilidae. In *Anopheles* it is present in Type I and III genes. In *Apis* it is only absent in TpnCIIa. Additional introns in other first EF-Hand locations have appeared in DmTpnCIa or in AmTpnCIIb genes.
- 3- The 4.21/1 intron is conserved in all TpnC genes, except those of *C. elegans*, where it is absent. Its length is close to 100 bp except in embryonic Type I genes, where it is bigger (up to 1.4 Kb; the precise intron sizes can be seen in Sup. Material Table B).
- 4- DmTpnCIIIb and AgTpnCIIIb1 have an additional intron just before the end of translation in position 4.28/1. Although this intron may not be present in AgTpnCIIIa and AgTpnCIIIb3, it is possible to find an alternative open reading frame that allows an intron position there. In general, this intron is Type III-specific, although it also occurs in the DpTpnCIb gene.
- 5- The first intron shows some variability, separating two of the three first codons. It does not occur in the two most closely related Type III *Anopheles* genes. Its size is quite variable (100 bp to 7 Kb).

The data on *C. elegans*, *A. mellifera*, *A. gambiae* and drosophilid genes indicate that intron positions are not as static as initially proposed (Yuasa and Takagi 2000; Yuasa and Takagi 2001). Insect variations are much more important than those found in other organisms. It is worth noticing that the three intron classes match with the functional types in Diptera, but not in *Apis*. For example, AmTpnCIIb is a Type II isoform with Class 3 intron positions (presence of intron 1.01/1, absence of 2.13/1). Also, Class 3 is characteristic of

Troponin C gene evolution in insects

dipteran flight-related genes (Type III) and Class 1 (presence of intron 2.13/1) marks the larval Type I genes plus the *Apis* Type III genes. This higher variation in intron position limits its utility in gene classification of distantly related species, but increases its applicability to analysing gene evolution in the more closely related protostome metazoan branch.

Insect Type III TpnC genes are clustered in the genome

Chromosome localizations for the orthologous genes are not conserved between *Drosophila* and *Anopheles* (see figure 3A). This is not a surprising result since this property is generally not very well conserved even within the *Anopheles* genus (Sharakhov et al. 2002). On the other hand, an intriguing feature is that the Type III genes appear tandemly clustered in all the insect genomes studied. The two adult Type III *Drosophila* TpnC genes are separated by 450 Kb in polytene region 41F in the second chromosome. In *Anopheles* there are four genes, with homology to the *Drosophila* Type III genes, clustered in a region of only 32 Kb in length (see figure 3B). The two first, named AgTpnCIIIb1 and IIIa, and the two last TpnC genes named AgTpnCIIIb2 and IIIb3 seem more closely related because they conserve their relative orientation in the genome and are of similar gene size. While this could indicate a parallel duplication of a IIIa and a IIIb precursor genes in *Anopheles*, three of the four Type III genes are paralogous to drosophilidae IIIb and one to IIIa gene as advanced in the names assigned based on sequence similarity.

TpnC genes are also chromosomally dispersed in *Apis mellifera*, although the precise genomic locations of the TpnC genes are still unknown. Nevertheless and similarly to the other insects counterparts, the two Type III genes are clustered, mapping together in a DNA sequence assembly of 5 Kb (figure 3C). In this respect, the larger distance separating the *D.*

Troponin C gene evolution in insects

melanogaster Type III genes may have arisen by insertion of additional sequences, favoured by their localization close to the heterochromatin in this species. This proposal could be additionally tested when other drosophilid genomes, such as *D. pseudoobscura*, are completely assembled. The relative orientation of these tandemly grouped genes does not correlate in the different insects so these duplications may have arisen independently in each insect order.

Conservation of TpnC functional features: gene expression patterns

RT-PCR experiments established the expression pattern of the five genes in *D. subobscura* and *D. virilis* (Figure 4) and confirmed the previously reported expression patterns in *D. melanogaster* (Herranz et al. 2004). The Type I group (TpnCIa and TpnCIb) is mainly expressed in embryonic and larval muscles (TpnCIa is also strongly expressed in the adult abdomens). The Type III group, formed by TpnCIIIa and the new found gene TpnCIIIb, are both expressed in adults, with IIIb restricted to the flight-related thoracic musculature, IFMs and TDTs. The fifth gene, TpnCII, shows a different expression pattern. This gene is expressed throughout *Drosophila* development (to some extent similarly to the larval TpnCIa) but in adults is expressed in heads and in the two main thoracic muscle groups (Herranz et al. 2004).

The same approach was applied to the analysis of TpnC expression patterns in honey bee developmental stages and body parts (Fig. 4, second row), using specific primer combinations for the five detected TpnC genes in the *Apis mellifera* genome. The AmTpnCIIIa and IIIb genes are maximally expressed in the thorax, with IIIa widely expressed in the adult and IIIb specifically in the thorax (IFM). The AmTpnCI and AmTpnCIIa genes show a general expression profile similar to DmTpnCIa (high expression

Troponin C gene evolution in insects

in larval stages) and DmTpnCII paralogous genes. The AmTpnCIIb gene (absent in Diptera) is not expressed at all in adults, showing a similar expression pattern to DmTpnCIb (absent in *Apis*).

Conservation of TpnC functional features: the repeated EF-Hand motifs

The main functional characteristics of the TpnC sequences are the repeated EF-Hand motifs and their divalent cation binding capabilities. Articles on calcium binding properties of protostome TpnCs have appeared (Allhouse et al. 1999; Ojima et al. 2000; Ueda et al. 2001, Qiu et al. 2003), including several *ecdysozoa* (crustaceans, arthropods, and a nematode, *Caenorhabditis elegans*), and also a *lophotrochozoan* (the mollusk scallop, *Chlamys nipponensis akazara*). These results indicate a functional role for the fourth EF-Hand in protostomes (low calcium affinity) very different from its high calcium affinity in mammals. In addition, in barnacle (*Balanus nubilus*, a crustacean) magnesium has been shown to bind a region overlapping the third EF-Hand in the TpnC carboxy terminal half, contributing to the stability of the TpnC-TnI interaction. Comparisons of the core sequences of the four EF-Hands of the insect TpnC genes show the conservation of the EF-Hand loop sequences (figure 5 and Supplementary material Fig D) and confirm the functional predictions of the genes in Drosophilidae, *Apis* and *Anopheles*. In several *ecdysozoa* (Allhouse et al. 1999; Ojima et al. 2000; Ueda et al. 2001; Qiu et al. 2003) and in all sequenced insect TpnCs the fourth EF-Hand is predicted to be functional. Most of the third EF-Hand sequences contain additional residues very similar to the equivalent *Balanus* sequence, consistent with an ability to bind magnesium ions and a role in the stabilisation of TpnC-TpnI interactions. An exception to this is the Type I isoforms in the Drosophilidae and AgTpnCIIIb2 (that have lost some key acidic positions). Modifications in this EF-Hand in

Troponin C gene evolution in insects

the three *Anopheles* Type IIIb genes, the main difference found among them, suggest a modulation in magnesium binding within these isoforms. The second EF-Hand is always functional in Type I and IIIa isoforms, but not in Type II proteins. Type IIIb proteins have putatively functional second EF-Hands in *Anopheles* and first EF-Hands in Drosophilidae, but AmTpnCIIIb can use neither of these EF-Hands for calcium binding. These findings strongly support the type-based classification of the insect TpnCs, but also identify special features of the Type III proteins, reinforcing the concept of a different origin of these isoforms in each insect group (see Discussion).

Based on direct Ca^{2+} binding assays, the TpnC EF-Hand functionality of the main adult isoforms in the hemipterous insect *Lethocerus* has been measured (Qiu et al. 2003). Taking all the information into account, LiTpnC3 is clearly a Type I orthologous gene. LiTpnC1, a Type III mainly thorax expressed isoform, has two functional EF-Hands (second and fourth), but LiTpnC4, also expressed in leg muscles, shows only one functional, the fourth, EF-Hand. The evolutionary distance between this hemimetabolous insect and the holometabolous ones is large, making the classification of LiTpnC4 difficult. Its sequence and expression pattern (in IFM muscles and in legs) suggest that it is a Type II (EF-Hands functionality and expression) or a Type III gene (sequence related to AmTpnCIIIa). Examination of its larval expression properties could resolve this question. Moreover, the existence of additional TpnCs in *Lethocerus* cannot be ruled out yet. It is possible that additional larval-specific genes have been missed since only adult data were presented by Qiu et al. (2003). Alternatively, since the hemiptera do not develop holometabolously, a gene, potentially related to a type II-III ancestor, may still be found in this insect order.

The TpnCIIIb isoforms in the Drosophilidae show an additional amino acid in the central alpha-helix of the dumbbell TpnC structure. This could have several consequences

Troponin C gene evolution in insects

for TpnC function such as the distance between N-terminal and C-terminal domains, the relative orientation, and the “flexibility” of the molecule (required for differential interactions in the troponin complex in presence or absence of the calcium signal). In *Anopheles* and *Apis*, other indels of residues occur in some of the more specialized genes (a general alignment appears as Supplementary Material). The characteristics of the calcium binding sites and the presence of an extra amino acid residue in the central alpha helix (summarized in Table II), may be important in defining how each TpnC functions.

DISCUSSION

Insects TpnCs can be classified in three groups based upon their functional properties.

In this work we have extended the type-based nomenclature of TpnC genes in *Drosophila melanogaster* to other insect TpnCs. Although in Diptera, all properties fit within this framework, the correlation decreases as we include further evolutionary separated groups in the analysis. Trees built with these sequences provide a useful representation of the relationship among gene types. The sequence similarity can be used as the starting point to classify new found genes in one of the types (Figure 1). On the other hand these trees are not always sufficient discriminating to establish the precise phylogeny of the gene family. At the base of the TpnC trees (grey ellipses), the significance (as measured by bootstrap values) drops to much lower values than in the peripheral branches (See supplementary material Fig A). The cDNA-based unrooted tree suggests a similar variation at the level of silent mutations (third base bias) in all genes. The lower variability of Type I and II proteins is probably related to functional constraints. Based on these considerations, the evolutionary history of these genes cannot be based solely on sequence variation, other properties have to be taken into account. The most difficult group to analyse in the trees is clearly the Type III

Troponin C gene evolution in insects

genes with large branches connected near the centre of the tree, suggesting, as discussed below, that these genes have evolved independently in different insect orders.

In general, the proposed classification agrees with other functional properties as EF-Hand putative functionality. The expression patterns strengthen the weight and usefulness of the classification. The exception is the *Apis mellifera* TpnCIIb gene, which is clearly a Type II gene by its DNA and protein sequences (including its EF-Hand properties), but shows a Type Ib expression profile. Since Type II genes have intermediate expression properties between Type I and Type III genes (they combine larvae and thoracic expression) and TpnCIIb gene is absent in *Apis*, it is not surprising that AmTpnCIIb shows a more restricted expression profile than expected. These results imply that this intermediate gene class can evolve more freely in different insect groups, being less determinate in expression than the Type I or Type III genes that are much more constrained into specific functions. The main strength of the type-based nomenclature is its generality. Once the sequence information is used to classify a new discovered gene, if the gene belongs to the Type I or III, we can predict the expression pattern of the gene. For intermediate Type II genes, individual properties may vary in each insect studied, to fill the expression profile gaps left by the other TpnC genes.

Reconstructing the evolutionary history of TpnC genes.

TpnC evolution has been well studied in vertebrates. The analysis of Kretsinger and Nakayama on EF-Hand protein evolution indicated that the vertebrate and the invertebrate TpnCs form two different groups (Nakayama and Kretsinger 1993; Nakayama et al. 1992). Our analysis indicates a very early separation of these groups, which apparently took place during the initial diversification of the late metazoan common ancestor (LMCA). TpnCs in

Troponin C gene evolution in insects

deuterostomes are more closely related to calmodulins than to protostome TpnCs, suggesting a polyphyletic origin for the TpnC genes. The high conservation of calmodulin paired with the plasticity of the multigene TpnC products may be also noted. Apart from this general consideration, an important increase in the number of TpnCs took place in insects and a minor expansion of TpnC gene number has also occurred in other *Ecdysozoa*, as seen in the nematode *C. elegans* genome and the available data in *Crustacea*. The availability of additional complete genome sequences in protostomes will possibly increase even more this variability.

To solve the discrepancies found between the gene types and the intron classes, a model for the TpnC gene evolution in insects is proposed (Figure 6). The model, integrating all information, is based on the concept that intron indels are quite rare events in evolution. The model resolves all the variability observed in TpnC genes intron positions as the minimum number of indel events, considering that single amino acid substitutions or insertions are much more easy to acquire. We argue that three TpnC precursors existed at the base of the holometabolous insect group diversification and each precursor gave rise at least to one TpnC gene in each organism. Our data strongly support this assumption, although obviously it is only one of the possible pathways for TpnC diversification. An alternative model would be to suppose that the intron indels have occurred secondarily, but we think it less likely for the reasons already discussed.

A single four EF hand polypeptide precursor may have existed in early metazoans, containing a common group of introns (Yuasa and Takagi, 2000, 2001). In *Lophotrochozoa* and nematodes, two TpnC isoforms appeared, in the case of nematodes certainly by gene duplication. During early insect diversification several duplications occurred. One TpnC gene copy diverged into the Class 1 TpnC precursor maintaining introns 1.01 and 2.13.

Troponin C gene evolution in insects

From the Class 2-3 precursor (from which intron 2.13 was lost), a new subdivision gave rise to the Class 2 (which also lost intron 1.01) and the Class 3 precursor. This must have happened sometime before *Apis* and dipterans diverged (more than 350 MY ago, Figure 6) and was probably associated with the transition to the holometabolous mode of development. Afterwards, the adult-specific genes suffered an expansion, close to the point at which the Diptera branched off (around 250 MY ago), evolving separately in *Apis*, Drosophilidae and *Anopheles*. In the case of *Apis* several duplications of the Class 1 precursor occurred evolving to produce a true Type I gene and two Type III genes. In the case of dipterans the expansion occurred in the Class 3 precursor branch, where a duplication produced the Type IIIa and Type IIIb genes. During insect speciation between 250 MY to 60 MY ago, the TpnC genes continued evolving reaching their present number in different insect orders.

It is known that duplicated genes evolve differentially in relation to their expression pattern (Hastings 1996; Ohta 2000), a property called subfunctionalisation (Force et al. 1999; Lynch and Force 2000; Lynch et al. 2001). The sequences of genes with broader expression patterns tend to be more highly conserved. In accordance with this, type I and II genes in the studied diptera insects are more conserved than Type III. The DmTpnCIb gene lost the 1.01/1 intron and gained more larval specificity (subfunctionalisation), while the DmTpnCIa gene substituted this intron for the one in the 1.30/1 position, and maintained a general muscle expression. The Type II gene in *Anopheles* acquired a new intron, 3.20/2. Our sequence data in Drosophilidae support a relatively recent duplication of the Type I TpnC genes in this genus. In *Apis* Type III genes sequences are better conserved than Type II gene sequences, reflecting both a multi-gene origin for Type II genes and also a different functionality since Type IIb (expressed during development and absent in adults) is much

Troponin C gene evolution in insects

more specialized than IIa (i.e. more conserved with respect to their drosophilidae paralogous genes). The expansion of the *Anopheles* Type III gene cluster is hard to date from the available information. Although their low conservation values suggest that they could be the result of ancient duplication events, the high specialization needs could be the reason for their present variability. Among Type III genes, the TpnCIIIa group is more highly conserved than the IIIb group, which is expressed only in the thoracic muscles. *Anopheles* IIIb genes evolved from a precursor incorporating the 3.13/1 intron by two consecutive, relative recent duplications, reflected in the expansion in size of that intron in the more distantly diverged gene (AgTpnCIIIb1) and the loss of the conserved intron just before the initiation codon in the other two genes.

Type III genes seem to be under stronger and faster evolutionary forces than the rest, exemplified in *Anopheles* by the different magnesium sensitivities of a set of isoforms needed to acquire a more specialized function. It is remarkable that the Type III insect genes are organized tandemly in clusters, especially in *Apis* where the Type III genes have a distinct origin from an intron Class 1 precursor yet are still clustered. This may provide a concerted transcriptional mechanism that allows these genes to be differentially modulated when that particular chromatin region becomes transcriptionally active. These possibilities point to a very specific fine-tuning role for the different Type III genes in insect flight musculature. The information obtained from the Drosophilidae argues that this situation has been maintained for the last 60 MY. Only small changes have occurred, including acquisition of a new stop codon mediated by an intron insertion (a newly appeared 4.28/1 intron), in the DpTpnCIb gene.

Flight acquisition as a driving force for TpnC diversification.

Troponin C gene evolution in insects

What were the evolutionary constraints in the insecta that were responsible for this considerable increase in gene number? The more specific driver selective forces behind TpnC gene diversification could have been the acquisition of the holometabolic pattern of development and flight. Our work suggests that a duplication of the insect TpnC ancestor occurred initially during the holometabolous transition while a second diversification of the adult isoforms (Type III), independently in different holometabolic insect orders, could have been driven by the acquisition of asynchronous flight muscles.

Adding to the complexity is the observation that more than one TpnC gene is expressed in each muscle type at varying levels, as described for dragonfly TpnT (Marden et al. 2001). Changes in EF-Hand properties and other features allow a large spectrum of specific combinations for each muscle type, such as IIIa plus IIIb in thoracic muscles, Ia and Ib (or Iib in *Apis*) for larval muscles, etc. The current analysis also suggests a correlation between wing beat frequency (related to the size of the adult animal) and the number of flight-related TpnC genes, from one in the big *Lethocerus*, two in Drosophilidae and *Apis* to four genes in the tiny *Anopheles*.

New Type III genes encoding for isoforms with slightly different structure and/or cation binding capabilities may have introduced new functional properties to the flight muscle. There is a general consensus that muscle asynchrony has appeared independently in insects several times during evolution (Pringle 1981; Dudley 2000; Josephson et al. 2000). This may explain why in hymenoptera (*Apis*) the Type III genes have evolved from a different gene precursor than in diptera, but converged to a similar protein complement to take account of the “new” asynchronous flight muscle function. The analysis of the intron indels in *Apis mellifera* Type III genes indicates that they may have arisen from a duplication of the Class 1 precursor (Figure 6). Sequence comparisons support this interpretation since

Troponin C gene evolution in insects

the *Apis* Type IIIa and IIIb genes appear in the same branch in the DNA tree close to its diffuse origin, while their protein sequences occur in the corresponding Type III subgroups (Figure 1 and Supplementary Material Fig A). A similar functional remodelling of the TpnC precursors is observed with AmTpnCIIB and its functional paralogous gene DmTpnCIb. This could be a clear case where functional diversification is also obtained by convergence starting from different ancestral genes.

In any case, evolution of muscle function including stretch-activation in insect species as in other organisms is undoubtedly more a consequence of comprehensive changes in the overall structure/function of the sarcomere than in the individual properties of particular components, such as TpnC. Changes in most components of the insect muscle sarcomeres are also known to have occurred. Further analyses taking into account not only the TpnC changes but the rest of the thick and thin filament components in insects and other organisms, could refine these proposals.

SUPPLEMENTARY MATERIAL

We present as on-line resources, Figure A (Phylogenetic relationships among insects and cDNA and protein-based TpnC phylogenetic trees), Figure B (the result of the screening of the *D. melanogaster*, *A. gambiae* and *C. elegans* genomes), Figure C (Protein and DNA Identity Matrices) and Figure D (general alignment of the *Drosophila*, *Anopheles* and *Apis* TpnCs in word and fasta formats). The unrooted trees in a Phylip format (files treeDNA and treePROT, showing bootstrapping values for Treeview application) and two additional Tables, A (Genbank accession numbers) and B (Intron positions in metazoans TpnCs) are provided.

ACKNOWLEDGMENTS

We want to express our thanks to the continuous and acknowledged efforts of consortia, especially as those currently in progress involving *Apis mellifera*, *Aedes aegypti* and, of course, *Drosophila pseudoobscura* genome sequencing. We are especially grateful to Drs. M. Cervera, R. Cripps, M. Manzanares and J. Sparrow for critical revisions as well as the anonymous referees of the Journal for many improving suggestions, Ms. Aida Villa for her technical support and Mr. Eulalio Castells Criado (Comunidad Autónoma de Madrid) and Mr. José Vicente Alonso (Industrias Alonso) for their help in providing the honey bee samples. This research was supported by grants from the Spanish Government PB96-0069, ESP1999-0379-C02 and PNE-008/2001-C. R. Herranz was supported during this work by a FPI fellowship of the Spanish Ministry of Science and Technology.

REFERENCES

- Allhouse LD, Guzman G, Miller T, Li Q, Potter JD, Ashley CC (1999) Characterisation of a mutant of barnacle TpnC lacking Ca²⁺-binding sites at positions II and IV. *Pflugers Arch* 438:30-9
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-402
- Baldauf SL (2003) Phylogeny for the faint of heart: a tutorial. *Trends Genet* 19:345-51
- Barbas JA, Galceran J, Torroja L, Prado A, Ferrus A (1993) Abnormal muscle development in the heldup3 mutant of *Drosophila melanogaster* is caused by a splicing defect affecting selected troponin I isoforms. *Mol Cell Biol* 13:1433-9
- Barton PJ, Cullen ME, Townsend PJ, Brand NJ, Mullen AJ, Norman DA, Bhavsar PK, Yacoub MH (1999) Close physical linkage of human troponin genes: organization, sequence, and expression of the locus encoding cardiac troponin I and slow skeletal troponin T. *Genomics* 57:102-9

Troponin C gene evolution in insects

- Benoist P, Mas JA, Marco R, Cervera M (1998) Differential muscle-type expression of the *Drosophila* troponin T gene. A 3-base pair microexon is involved in visceral and adult hypodermic muscle specification. *J Biol Chem* 273:7538-46
- Bernstein SI, O'Donnell PT, Cripps RM (1993) Molecular genetic analysis of muscle development, structure, and function in *Drosophila*. *Int Rev Cytol* 143:63-152
- Dudley R (2000) *The Biomechanics of Insect Flight: Form, Function, Evolution*. Princeton University Press.
- Filatov VL, Katrukha AG, Bulargina TV, Gusev NB (1999) Troponin: structure, properties, and mechanism of functioning. *Biochemistry (Mosc)* 64:969-85
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-45
- Fyrberg C, Parker H, Hutchison B, Fyrberg E (1994) *Drosophila melanogaster* genes encoding three troponin-C isoforms and a calmodulin-related protein. *Biochem Genet* 32:119-35
- Fyrberg EA, Mahaffey JW, Bond BJ, Davidson N (1983) Transcripts of the six *Drosophila* actin genes accumulate in a stage- and tissue-specific manner. *Cell* 33:115-23
- Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19:748-61
- Gordon AM, Homsher E, Regnier M (2000) Regulation of contraction in striated muscle. *Physiol Rev* 80:853-924
- Grabarek Z, Tao T, Gergely J (1992) Molecular mechanism of troponin-C function. *J Muscle Res Cell Motil* 13:383-93
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714-23
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser* 41:95-98
- Hastings KE (1996) Strong evolutionary conservation of broadly expressed protein isoforms in the troponin I gene family and other vertebrate gene families. *J Mol Evol* 42:631-40

Troponin C gene evolution in insects

- Herranz R, Díaz-Castillo C, Nguyen TG, Lovato TL, Cripps RM, Marco R (2004) Expression pattern characterization of the whole TpnC gene repertoire during *Drosophila* development. *Gene Expr Patterns* 4:183-190
- Herzberg O, James MN (1985) Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. *Nature* 313:653-9
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298:129-49
- Huang QQ, Jin JP (1999) Preserved close linkage between the genes encoding troponin I and troponin T, reflecting an evolution of adapter proteins coupling the Ca²⁺ signaling of contractility. *J Mol Evol* 49:780-8
- Hulo N, Sigrist CJ, Le Saux V, Langendijk-Genevaux PS, Bordoli L, Gattiker A, De Castro E, Bucher P, Bairoch A (2004) Recent improvements to the PROSITE database. *Nucleic Acids Res* 32 Database issue:D134-7
- Josephson RK, Malamud JG, Stokes DR (2000) Asynchronous muscle: a primer. *J Exp Biol* 203 Pt 18:2713-22
- Kawasaki H, Nakayama S, Kretsinger RH (1998) Classification and evolution of EF-Hand proteins. *Biometals* 11:277-95
- Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459-73
- Lynch M, O'Hely M, Walsh B, Force A (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789-804
- Marden JH, Fitzhugh GH, Girgenrath M, Wolf MR, Girgenrath S (2001) Alternative splicing, muscle contraction and intraspecific variation: associations between troponin T transcripts, Ca²⁺ sensitivity and the force and power output of dragonfly flight muscles during oscillatory contraction. *J Exp Biol* 204:3457-70
- Marsden BJ, Shaw GS, Sykes BD (1990) Calcium binding proteins. Elucidating the contributions to calcium affinity from an analysis of species variants and peptide fragments. *Biochem Cell Biol* 68:587-601

Troponin C gene evolution in insects

- Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* 11:699-704
- Nakayama S, Kretsinger RH (1993) Evolution of EF-Hand calcium-modulated proteins. III. Exon sequences confirm most dendrograms based on protein sequences: calmodulin dendrograms show significant lack of parallelism. *J Mol Evol* 36:458-76
- Nakayama S, Moncrief ND, Kretsinger RH (1992) Evolution of EF-Hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories. *J Mol Evol* 34:416-48
- Ohno S (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. *Semin. Cell Dev. Biol.* 10:517-522
- Ohta T (2000) Evolution of gene families. *Gene* 259:45-52
- Ojima T, Koizumi N, Ueyama K, Inoue A, Nishita K (2000) Functional role of Ca(2+)-binding site IV of scallop TpnC. *J Biochem (Tokyo)* 128:803-9
- Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357-8
- Powell JR (1997) Progress and prospects in Evolutionary Genetics: The Drosophila Model. New York Oxford University press, NY
- Pringle JW (1981) The evolution of fibrillar muscle in insects. *J. Exp. Biol.* 94:1-14
- Qiu F, Lakey A, Agianian B, Hutchings A, Butcher GW, Labeit S, Leonard K, Bullard B (2003) TpnC in different insect muscle types: identification of two isoforms in *Lethocerus*, *Drosophila* and *Anopheles* that are specific to asynchronous flight muscle in the adult insect. *Biochem J* 371:811-21
- Reinach FC, Farah CS, Monteiro PB, Malnic B (1997) Structural interactions responsible for the assembly of the Troponin Complex on the muscle thin filament. *Cell Struct Funct* 22:219-23
- Rokas A, Holland PW (2000) Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* 15:454-459
- Schwede T, Diemand A, Guex N, Peitsch MC (2000) Protein structure computing in the genomic era. *Res Microbiol* 151:107-12

Troponin C gene evolution in insects

- Sharakhov IV, Serazin AC, Grushko OG, Dana A, Lobo N, Hillenmeyer ME, Westerman R, Romero-Severson J, Costantini C, Sagnon N, Collins FH, Besansky NJ (2002) Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* 298:182-5
- da Silva AC, Kendrick-Jones J, Reinach FC (1995) Determinants of ion specificity on EF-Hands sites. Conversion of the $\text{Ca}^{2+}/\text{Mg}^{2+}$ site of smooth muscle myosin regulatory light chain into a Ca^{2+} -specific site. *J Biol Chem* 270:6773-8
- Sparrow JC (1995) Muscle. Flight and phosphorylation. *Nature* 374:592-3
- Terami H, Williams BD, Kitamura S, Sakube Y, Matsumoto S, Doi S, Obinata T, Kagawa H (1999) Genomic organization, expression, and analysis of the TpnC gene pat-10 of *Caenorhabditis elegans*. *J Cell Biol* 146:193-202
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-80
- Ueda T, Katsuzaki H, Terami H, Ohtsuka H, Kagawa H, Murase T, Kajiwarra Y, Yoshioka O, Iio T (2001) Calcium-bindings of wild type and mutant TpnCs of *Caenorhabditis elegans*. *Biochim Biophys Acta* 1548:220-8
- Vassilyev DG, Takeda S, Wakatsuki S, Maeda K, Maeda Y (1998) The crystal structure of TpnC in complex with N-terminal fragment of troponin I. The mechanism of how the inhibitory action of troponin I is released by Ca^{2+} -binding to TpnC. *Adv Exp Med Biol* 453:157-67
- Venkatesh B, Erdmann MV, Brenner S (2001) Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc Natl Acad Sci U S A* 98:11382-7
- Yuasa HJ, Cox JA, Takagi T (1998) Diversity of the TpnC genes during chordate evolution. *J Biochem (Tokyo)* 123:1180-90
- Yuasa HJ, Takagi T (2000) The genomic structure of the scallop, *Patinopecten yessoensis*, TpnC gene: a hypothesis for the evolution of TpnC. *Gene* 245:275-81
- Yuasa HJ, Takagi T (2001) Genomic structure of the sandworm, *Perinereis vancaurica* tetradentata, TpnC. *Gene* 268:17-22

FIGURE LEGENDS

Figure 1. TpnC protein-based tree in insects. Identity matrices (Supplementary material Fig C) were constructed using the BIOEDIT® (Hall 1999) program after alignment with ClustalW algorithm (Thompson et al. 1994) followed by manual refinement (Baldauf 2003). Phylogenetic trees were obtained from these alignments using “Fitch-Margoliash and Least-Squares Distance Methods” algorithm and Phylodendron version 0.8d, beta 1999 & ClustalW v1.75 (Neighbour-Joining method with 1000 bootstraps). Tree output was performed with TreeView® v1.6.6 (Page 1996). Branches with bootstrap values below 70% have been collapsed and shaded within a grey ellipse. Bootstrap values can be seen in the trees included as supplementary material. The sequences can be grouped in three types, types I and II, clearly clustered, and type III, including the rest. Following the *D. melanogaster* analysis, we propose here that insect TpnC genes should be named first by its type (I, II or III) followed by a letter if more than one gene exists in the same group. Type I and Type II groups are clearly separated phylogenetically. Type III sequences are much more variable and can be divided in IIIa and IIIb subgroups using the protein tree. *Apis* type III TpnCs are so distantly related with their dipteran counterparts that they appear together in the IIIa subgroup of the cDNA tree (bootstrap value 73%, see supplementary material Fig A), suggesting an independent diversification of this group in insects. These gene types based on sequence similarity have been used to propose a new comprehensive nomenclature for all insect TpnC genes. They are also useful to classify additional phenotypic features such as the EF hands properties and the gene expression patterns. The scale bar indicates the branch length distances as numbers of substitutions per site.

Figure 2. Structure of the predicted TpnC genes in insects. On the left are scale drawings of TpnC gene structure in different species. Non-coding exons are shown as white boxes and coding regions are denoted by black boxes. On the right are diagrams representing the intron positions of TpnC genes. Introns are defined by several numbers following the Nakayama and Kretsinger annotation (1993). The first number indicates the EF hand interrupted by the intron, sequentially numbered from N- to C-terminus. The second number (following the period) shows the number of the amino acid within the EF hand where the intron is located. The third number (following the slash) indicates its phase: 0 means the intron lies between codons, 1 means between first and second nucleotides, and 2 means between the second and the third. Grey circles represent conserved introns, white circles represent gene specific introns and black circles are class defining introns. For comprehensiveness, the numbers inside some white circles indicate the amino acid interrupted by the corresponding gene specific intron when more than one has been found in a similar position as can be seen at the bottom. Gene structures are well preserved in Drosophilidae (Supplementary Table B). *Anopheles* TpnC genes show bigger introns than their drosophilid counterparts, but intron positions are almost completely preserved among each gene type. In contrast, intron positions show an apparently higher degree of variation between *Apis* and Diptera TpnCs. Three kinds of organization based on the intron positions can be distinguished: Class 1 contains intron 2.13 and 1.01 (except in Drosophilidae), Class 3 maintains only the 1.01 intron and Class 2 lacks both introns. Thus, intron class and gene type totally match for dipterans but not in the case of *Apis*.

Figure 3. Genomic features of Type III TpnC genes. **A)** Comparison of the chromosomal locations of the TpnC genes in *D. melanogaster* and *A. gambiae* genomes. Their relative locations, including the pericentromeric position of the Type III cluster in *D. melanogaster*, are not conserved. **B)** The cluster of Type III genes in the *Anopheles gambiae* genome clone AAAB0108960 is located cytologically in the 25D position in the middle of 2L chromosome arm. Although both AgTpnCIIIb1 and AgTpnCIIIa genes conserve the same orientation in the chromosome, only the second one is a DmTpnCIIIa orthologue. The other three genes are paralogues of the DmTpnCIIIb gene. AgTpnCIIIb2 and AgTpnCIIIb3 genes are separated by only 500 bp. **C)** Type III genes are also clustered in *Apis mellifera* genome. There are separated by only 1500 bp and the cluster size is 5085 bp.

Figure 4. Expression pattern of TpnCs during development and in adult body parts.

Upper panel: RT-PCR experiments performed in two drosophilid species reveal conservation in the expression profile of TpnC genes in this genus. RNA was isolated from embryos (E), Early and Late Larvae (EL, LL), Early and Late Pupae (EP, LP), and adult body parts (Head-H, Thorax-Tx and Abdomen-Ab) and RT-PCR performed with specific probes for each TpnC gene (see Materials and Methods). Type I genes are predominantly expressed at larval stages (Ib) and in adult abdomens (Ia). Type II genes are also expressed throughout development, and show particularly high expression levels in adult heads. Type III gene expression is turned on in pupae reaching maximum levels in thoraces (exclusively for TpnCIIIb).

Bottom panel: Expression profile of TpnCs during *Apis* development. RNA was isolated from unclosed larvae (mid-larval stage, ML), closed larvae (late larval stage, LL), Early (non-pigmented) and Late (fully pigmented) Pupae (EP, LP), adult body parts (Head-H, Thorax-Tx and Abdomen-Ab), thorax isolated fibers (DV-IFM, DL-IFM), thoracic tubular muscles (Other) and legs. RT-PCR was performed in the presence (+) and absence (-) of reverse transcriptase using primer pairs specific for each *Apis* TpnC gene (see Materials and Methods). All *Apis* genes show an expression pattern similar to their relatives in Drosophilidae with the exception of the AmTpnCIIb gene that is not expressed in any part of the mature honey bee thus resembling the drosophilid TpnCIb gene expression profile.

Figure 5. Comparison of the core sequences of the TpnC EF-Hands in *Drosophila*, *Anopheles* and *Apis*. EF-Hand properties were predicted by conservation of the canonical EF-Hand sequence on the Calcium Binding Protein data library (Marsden et al. 1990) and with arthropods EF-Hands with known calcium or magnesium binding properties. The same predictions are obtained using the Prosite software (Hulo et al. 2004). In grey shading and black letters the predicted low affinity Ca^{2+} binding (regulatory) functional sites and in grey shading and white letters the potentially Mg^{2+} binding EF-Hands. Non-functional EF-Hands are not highlighted. Although they are not actually identical, the *D. melanogaster* (*D.m.*) TpnC sequences are highly conserved in *D. subobscura* (*D.s.*) and *D. virilis* (*D.v.*) (plus *D. silvestris* (*D.si*) for DsiTpnCIIIb). All *D. pseudoobscura* EF hands (not shown) are conserved in *D. subobscura*. The sequences included in boxes separate the three TpnC gene groups (Type I in the upper box, Type III in the middle and Type II in the lower box) in different species. If amino acids are conserved a dot appears in the following sequence(s) using the DmTpnCIa sequence as reference. As also shown in Table II, *Lethocerus indicus* (Li) TpnC4 is at the limit of Type II-III separation, this gene has only a single functional EF-Hand (Type II), however, the two first EF-Hands mostly resemble those of AmTpnCIIIa.

Figure 6. A model for the evolutionary diversification of TpnC genes in insects. This scheme integrates all the available information, namely, expression patterns, exon-intron features, EF-Hand conservation and general protein and cDNA sequence comparisons of the TpnC genes in *Apis*, *Anopheles* and Drosophilidae. Duplication events (squares), addition of residues and intron gain or loss are indicated. Different colors show the species specificity of the branches (green for Drosophilidae, red for *Anopheles* and blue for *Apis*). Gene names appear in different styles (plain text for type I, italics for *Type II* and bold text for **Type III** genes). Diagrammatic schemes of the proteins appear at the right and left sides of the figure showing the four EF hand sites (red and blue dots indicate Mg^{2+} and Ca^{2+} binding ability respectively). Intron positions in them are marked by arrow heads. The expression patterns of the Drosophilidae and *Apis* genes are also indicated. After divergence of the three insect TpnC gene precursors, intron positions and EF-Hand capabilities changed, likely leading to more specialized functions. Type II genes accumulated a relatively low number of variations, and their higher sequence conservation may result from their broader expression than Type III genes. Class I in *Apis* and Class III genes in diptera suffered up to three independent duplication events linked to the flight related Type III genes, especially exemplified in *Anopheles*, to obtain the Type IIIa and a set of up to three Type IIIb isoforms, likely able to promote a “fine-tuning” function in thoracic flight-related muscles.

Troponin C gene evolution in insects

	<i>Drosophila</i>	<i>Anopheles</i>	<i>Apis</i>	<i>Caenorhabditis</i>	<i>Fugu</i>	Mammals
TpnT	1*	1*	1*	1*	3	3 SS*, FS*, C*
TpnI	1*	1*	1*	1*	3	3 SS*, FS*, C
TpnC	5 73F – Ia 47D – Ib 25D – II 41C – IIIa 41F – IIIb	6 AgCG55381 – I AgCG47362 – II AgCG51999# – IIIa AgCG51999# – IIIb1 AgCG52004 – IIIb2 AgCG52005 – IIIb3	5 AmTpnCI AmTpnCIIa AmTpnCIIb AmTpnCIIIa AmTpnCIIIb	2 TpnC1-pat10 TpnC2	4 Sk1, Sk2 C1, C2	2 FS, SS/C

TABLE I: Gene number and names of the Troponin Complex components in different species.

The asterisk (*) means that the gene is differentially spliced into multiple isoforms. SS=Slow Skeletal, FS=Fast Skeletal & C=Cardiac Isoform. The known vertebrate TpnC gene sequences are also listed, including the results of a search in *Fugu rubripes* (teleost fish) sequence data base using mammalian TpnC genes as queries. The polytene cytological localization originally proposed to distinguish the *D. melanogaster* genes (Fyrberg et al, 1994) and the Celera Genomics numbers of *Anopheles* genes are shown together with the type-based nomenclature that we have introduced. The # indicates that type IIIa and IIIb1 genes are included as a single, alternatively spliced, gene in the Celera annotation, AgCG51999.

Troponin C gene evolution in insects

Gene Product	Type	First EF-Hand	Second EF-Hand	Central Indels	Third EF-Hand	Fourth EF-Hand
DmTpnCIa	I	-	+	-	-	+
DmTpnCIb	I	-	+	-	-	+
AgTpnCI	I	-	+	-	*	+
AmTpnCI	I	-	+	-	*	+
LiTpnC3	I	-	+	-	*	+
DmTpnCII	II	-	-	-	*	+
AgTpnCII	II	-	-	+Q	*	+
AmTpnCIIa	II	-	-	-	*	+
AmTpnCIIb	II	-	-	+Q-LD	*	+
LiTpnC4	II-III	-	-	+G	*	+
DmTpnCIIIa	III	-	+	-	*	+
AgTpnCIIIa	III	-	+	-	*	+
AmTpnCIIIa	III	-	+	-	*	+
LiTpnC1	III	-	+	+DNE	*	+
DmTpnCIIIb	III	+	-	+E	*	+
AgTpnCIIIb1	III	-	+	+E	*	+
AgTpnCIIIb2	III	-	+	+E	-	+
AgTpnCIIIb3	III	-	+	+E	*	+
AmTpnCIIIb	III	-	-	+E	*	+

TABLE II. Conserved features of different insect TpnC proteins. Calcium binding (+) or not (-), magnesium binding (*) in EF-Hand domains are indicated. The amino acid indels in the central TpnC helix are also shown. Several 3D predictions were obtained remotely from the Expasy server using the “first approach method” with different probe combinations and analyzed with SwissPdbViewer (Guex and Peitsch 1997; Schwede et al. 2000), suggesting that these indels could affect the TpnC 3D structures. In addition to the data generated in *Drosophila*, *Apis* and *Anopheles* we have introduced the results of *Lethocerus* (giant water bug, Qiu et al, 2003).

FIGURE 1

FIGURE 2

FIGURE 3

FIGURE 4

FIGURE 5

Mammals Consensus	D	*	D	G	D/S	G	#	I	D/G/S	*	D/E	E
Seq. core: 12aa	98%	--	73%	51%	56/22	89%	--	70%	33/14/19	--	27/32	86%

<u>Loop 1st EF-Hand</u>		<u>Loop 2nd EF-Hand</u>	
DmTpnCIa	DHQKTGSIPTM	DmTpnCIa	DEDKSGRLEFGE
DsTpnCIa	••••S•••••	DmTpnCIb	••••••••E•
DmTpnCIb	••••••••••	DvTpnCIb	••SKC••••E•
AgTpnCI	•RE••••SSD•	AgTpnCI	•••••QI•A•
AmTpnCI	•RE•S••••D•	AmTpnCI	•A•••••E•
LiTnC3	•RERS••••D•	LiTnC3	•A•••••E•
DmTpnCIIa	•PE•N•Y•N•A•	DmTpnCIIa	••G••QI•E•
AgTpnCIIb1	•KE•••••DV	AgTpnCIIb1	••E••Q••E•
AgTpnCIIa	•QE•K•C•G•Q•	AgTpnCIIa	•A•G••E••E•
AgTpnCIIb2	•IE•K••SV•V	AgTpnCIIb2	•V•E••QI•D•
AgTpnCIIb3	•IE•K••SL•V	AgTpnCIIb3	••E••QI•E•
AmTpnCIIa	•••K••G•D•	AmTpnCIIa	••G••Q••E•
LiTnC1	AQE•S•Y•EAN•	LiTnC1	•A•G•Q••E•
AmTpnCIIb	•PD•K•T•T•DT	AmTpnCIIb	•PFG••EMN•E•
LiTnC4	TPEGS•RVS•DQ	LiTnC4	•PFGN•DID•DS
DmTpnCIIb	•DGA••EHAD	DmTpnCIIb	•KGTT•K•D•SQ
DsTpnCIIb	•DGA••DHTD	DsTpnCIIb	•KGTS•K•D•GQ
DsiTpnCIIb	•LDGS••DHSD	DsiTpnCIIb	•KGTSNK•N•AQ
DvTpnCIIb	•LEG••W•DHKD	DvTpnCIIb	•KGTTEK•N•SQ
DmTpnCII	•T•••F•E•LR	DmTpnCII	•PEDT•KVN•DG
DvTpnCII	•T••S•F•E•LR	AgTpnCII	•PEST••VN•DG
AgTpnCII	•T•••F•E•IK	AmTpnCIIa	•PEGT•KVN•DG
AmTpnCIIa	•TT•S•F•D•LK	AmTpnCIIb	•EEG••K•N•DS
AmTpnCIIb	•ST•S•R•EK•K		

<u>Loop 3rd EF-Hand</u>		<u>Loop 4th EF-Hand</u>	
DmTpnCIa	DKQGNFIPTTC	DmTpnCIa	DSDGSGTVDFDE
DmTpnCIb	•••••Y••S•	DmTpnCIb	••••••••••
AgTpnCI	•E••••••••	AgTpnCI	••••••••••
AmTpnCI	•E••Y•••••	AmTpnCI	••••••••••
LiTnC3	•E••Y•P•S•	LiTnC3	•T••••••••
DmTpnCIIa	•E••Y•T•GV	DmTpnCIIa	••••••••••
AgTpnCIIb1	•DAK•YL•VEE	AgTpnCIIb1	•A•••••E•
AgTpnCIIa	•E••Y•T•QV	AgTpnCIIa	••••••••••
AgTpnCIIb2	•N•T•••LDV	AgTpnCIIb2	•A•••••E•
AgTpnCIIb3	•D•T•••VNS	AgTpnCIIb3	•A•S••••E•
AmTpnCIIa	•E••Y•T•AV	AmTpnCIIa	•A••••L•••
LiTnC1	•D••Y•T•EV	LiTnC1	•S••••••••
AmTpnCIIb	•E••Y•T•DV	AmTpnCIIb	•A•••••E•
LiTnC4	•E••Y•S•DV	LiTnC4	•A•••••E•
DmTpnCIIb	•E•K•YLT VAT	DmTpnCIIb	•A••••••••
DmTpnCII	•RE••Y•T•ST	DmTpnCII	•T••••••••
AgTpnCII	•RE••Y•T•ST	AgTpnCII	•T••••••••
AmTpnCIIa	•RE••Y•T•AT	AmTpnCIIa	•T••••••••
AmTpnCIIb	•RE••Y••SS	AmTpnCIIb	•T••••••••

FIGURE 6