

# An Evolutionary Analysis of Flightin Reveals a Conserved Motif Unique and Widespread in Pancrustacea

Felipe N. Soto-Adames · Pedro Alvarez-Ortiz ·  
Jim O. Vigoreaux

Received: 28 June 2013 / Accepted: 11 November 2013 / Published online: 23 November 2013  
© Springer Science+Business Media New York 2013

**Abstract** Flightin is a thick filament protein that in *Drosophila melanogaster* is uniquely expressed in the asynchronous, indirect flight muscles (IFM). Flightin is required for the structure and function of the IFM and is indispensable for flight in *Drosophila*. Given the importance of flight acquisition in the evolutionary history of insects, here we study the phylogeny and distribution of flightin. Flightin was identified in 69 species of hexapods in classes Collembola (springtails), Protura, Diplura, and insect orders Thysanura (silverfish), Dictyoptera (roaches), Orthoptera (grasshoppers), Phthiraptera (lice), Hemiptera (true bugs), Coleoptera (beetles), Neuroptera (green lacewing), Hymenoptera (bees, ants, and wasps), Lepidoptera (moths), and Diptera (flies and mosquitoes). Flightin was also found in 14 species of crustaceans in orders Anostraca (water flea), Cladocera (brine shrimp), Isopoda (pill bugs), Amphipoda (scuds, sideswimmers), and Decapoda (lobsters, crabs, and shrimps). Flightin was not identified in representatives of chelicerates, myriapods, or any species outside Pancrustacea (Tetraconata, sensu Dohle). Alignment of amino acid

sequences revealed a conserved region of 52 amino acids, referred herein as WYR, that is bound by strictly conserved tryptophan (W) and arginine (R) and an intervening sequence with a high content of tyrosines (Y). This motif has no homologs in GenBank or PROSITE and is unique to flightin and paraflightin, a putative flightin paralog identified in decapods. A third motif of unclear affinities to pancrustacean WYR was observed in chelicerates. Phylogenetic analysis of amino acid sequences of the conserved motif suggests that paraflightin originated before the divergence of amphipods, isopods, and decapods. We conclude that flightin originated de novo in the ancestor of Pancrustacea > 500 MYA, well before the divergence of insects (~400 MYA) and the origin of flight (~325 MYA), and that its IFM-specific function in *Drosophila* is a more recent adaptation. Furthermore, we propose that WYR represents a novel myosin coiled-coil binding motif.

**Keywords** Insect flight muscle · Flightin · *Drosophila* · Hexapods · Crustaceans

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-013-9597-5) contains supplementary material, which is available to authorized users.

F. N. Soto-Adames · P. Alvarez-Ortiz · J. O. Vigoreaux (✉)  
Department of Biology, University of Vermont, 109 Carrigan  
Drive, Burlington, VT 05405, USA  
e-mail: jvigorea@uvm.edu

*Present Address:*

F. N. Soto-Adames  
Illinois Natural History Survey, University of Illinois,  
Champaign, IL 61820, USA

J. O. Vigoreaux  
Department of Molecular Physiology and Biophysics, University  
of Vermont, Burlington, VT 05405, USA

## Introduction

Among eukaryotes, Insecta is the most diverse class on Earth and widely considered the most successful terrestrial metazoan. Approximately 900,000 species of insects are known, but estimates for the number of insect species range from 2,000,000 to 30,000,000. Insect success, and particularly winged insects are measured not only in terms of sheer number of species, but also in terms of distribution, colonization of diverse habitats, and population sizes. Insects have a 400 million year history and today account for nearly 60 % of all known species of living organisms

and close to 80 % of all animal species (Wilson 1992; Grimaldi and Engel 2005). A great radiation of modern insects took place during the Carboniferous and Permian and several factors are likely to have played an important role in their diversification and evolutionary success. Chief among these is the origin of flight (~325 MYA), considered a pivotal event in the history of life as the transition to air opened up vast opportunities for colonization of new habitats and niche exploitation and subsequent adaptive radiation (Brodsky 1994; Dudley 2000). The advantages of flight are still evident to date inasmuch as more than 90 % of known hexapod species are winged insects (Grimaldi and Engel 2005).

Another factor that is considered to have contributed significantly to the diversification of winged insects is miniaturization. High wing beat frequency (>100 Hz), typically associated with asynchronous muscles is likely to have played a key role in the diminution of insect body size (Dudley 2000). For all insects, wing beat frequency scales with body mass<sup>-0.24</sup> (Dudley 2000) suggesting that factors that influence wing beat frequency may have experienced strong selection pressure in the historical push toward miniaturization. The flight system of small insects is essentially a resonance system whose fundamental frequency is influenced by the physical properties of its components (i.e., wings, cuticle, and flight muscles). Factors that influence muscle stiffness are of particular interest given their proportional effect on wing beat frequency (Hyatt and Maughan 1994). High muscle stiffness could also be an indirect consequence of the selection for minimization of metabolic cost for contraction (Dudley 2000), i.e., the use of elastic energy storage as a strategy to minimize metabolic power expenditure. Thus, understanding the evolution of flight muscle requires an examination of the factors that influence muscle stiffness and wing beat frequency.

Asynchronous muscles are activated by stretch and characteristically show a high resting stiffness, the magnitude of which is proportional to the amplitude of stretch activation (Moore 2006). In *Drosophila*, two orthogonal and antagonistic asynchronous muscles, the dorsoventral muscles, and the dorsolongitudinal muscles (together the indirect flight muscles, or IFM) power flight through reciprocal stretch activation: contraction of the dorsoventral muscles stretches and initiates the contraction of the dorsolongitudinal muscles, which in turn contract and initiate the contraction of the dorsoventral muscles. Stretch activated insect flight muscles operate at low strains [e.g., 2–5 % for *Drosophila* IFM; Chan and Dickinson 1996], a condition mandated by the nearly complete overlap of thick and thin filaments (i.e., narrow I bands). The combination of these two factors, high stiffness and low strain (i.e., change in muscle length divided by original length), is

probably interrelated and necessary for efficient force transmission conducive to oscillatory work output. Thus, an important step in flight muscle evolution was the acquisition of structural elements that increase stiffness of the myofibril (Maughan and Vigoreaux 2004). Key among these is flightin, an IFM-specific protein known to have a direct effect on the stretch activation response (Vigoreaux et al. 1993; Henkin et al. 2004). Studies of flightin can therefore provide a mechanistic link between protein biophysics and physiological adaptations in muscle and the selective agents responsible for the evolutionary diversification of insects.

Flightin is a 20,000 Da protein first identified in *Drosophila melanogaster* IFM (Vigoreaux et al. 1993). Studies have shown that flightin is a multiply phosphorylated, myosin rod binding protein that is essential for the normal assembly and stiffness of thick filament, sarcomere stability, and contractile activity of *Drosophila* IFM (Vigoreaux 1994; Vigoreaux et al. 1998; Reedy et al. 2000; Ayer and Vigoreaux 2003; Henkin et al. 2004; Barton et al. 2005; Contompasis et al. 2010). *Drosophila* with a null mutation in the flightin gene (*fn*<sup>0</sup>) are viable, but flightless due to age-dependent atrophy of their flight musculature. A characteristic feature of *fn*<sup>0</sup> IFM is that thick filaments and sarcomeres are, on average, ~30 % longer than thick filaments and sarcomeres in wild-type IFM at the late pupal stage (Reedy et al. 2000). Despite their otherwise normal appearance, *fn*<sup>0</sup> sarcomeres are structurally compromised as evidenced by their inability to withstand contractile forces, resulting in sarcomere breakdown and fiber hypercontraction shortly after eclosion (Reedy et al. 2000; Nongthomba et al. 2003). The structural and functional flight muscle defects in *fn*<sup>0</sup> are fully rescued with the introduction of a normal *fn*<sup>+</sup> transgene, demonstrating that the main function of flightin is in the IFM (Barton et al. 2005). This conclusion is consistent with high throughput and tissue expression data that show flightin expression limited to pupal and adult stages and excluded from adult tissues other than the IFM (flybase.org). In addition to *Drosophila*, flightin has been identified in the waterbug, *Lethocerus indicus*, where it also has been reported to be expressed exclusively in the IFM (Ferguson et al. 1994; Qiu et al. 2005).

The co-occurrence of flightin and asynchronous flight muscle in *Drosophila* and *Lethocerus* raises the interesting possibility that the origin of the unique muscle physiology might correlate with the origin of flightin or changes in flightin expression patterns. Despite the vital importance of flightin for IFM function in *Drosophila*, and its potentially important role in the context of asynchronous muscle evolution, there is no information about its phylogenetic distribution. Here, we present the results of a systematic survey of Genbank for the presence of flightin among

metazoans. We show that flightin is characterized by a conserved ~50 amino acid motif located near the middle of the protein sequence. Our analysis shows that flightin is encoded by a single gene present in all species of crustaceans and hexapods examined, which range from basal branchiopods to higher order insects, suggesting that flightin is common throughout Pancrustacea [=Tetraconata, (Dohle 2001; Richter 2002)]. In contrast, flightin was not found in other major metazoan groups, including other arthropod sub-phyla, annelids, and vertebrates. We also report the presence of a putative paralogue of flightin, named paraflightin with which flightin shares many of the amino acids in the conserved motif.

## Results

### Taxonomic Distribution of Flightin

Flightin was identified in 50 species of hexapods (49 insects and one springtail) and 9 species of crustaceans for which the complete open reading frame was available (Supplementary Table 1). These include sequences from complete genomes, protein RefSeq and EST libraries. Alignment of these sequences revealed a highly conserved region extending from amino acids 84 through 135 (from herein, numbers refer to positions in the sequence of *D. melanogaster*; Supplementary Fig. 1). Sequences on the N- and C-termini to this region are highly variable and no consensus alignment was recovered in comparisons across orders. We then searched the EST library limiting the subject to the highly conserved region and retrieved flightin in more than 40 additional species. In total, we identified flightin in 69 species of hexapods and 14 species of crustaceans (Supplementary Table 1). The range of species in which the flightin sequence was identified includes insect with asynchronous flight muscles (flies, bees, wasps, ants, mosquitoes, beetles, water bugs, aphids, and leafhoppers) and synchronous flight muscles (silk and turnip moths, migratory locust, cockroaches, and termites), and species without flight muscles or wings (silverfish, bristletail, springtail, proturan, dipluran, and all crustacea). All of these species fall within the Pancrustacea (Tetraconata) clade.

Queries on other arthropods using the *Drosophila* sequence were unsuccessful, despite using different search strategies. Only with thresholds of  $10^{-3}$  or lower when searching individual libraries were possible matches detected in some chelicerates (e.g., horseshoe crab, deer and brown ear ticks, tarantula, scorpions, chiggers, to name a few). These sequences contain at least five of the conserved amino acids found in the sequences from the pancrustacean clade. However, in all instances the matching sequences appear to be part of very different open reading

frames (see section on Paraflightin). No matches were detected among the myriapods examined.

A search of Protostomia, excluding hexapods and crustaceans, using the conserved sequence from the bristletail *Tricholepisma* as query, resulted in a single hit to the medicinal leech (*Hirudo medicinalis*). However, no putative flightin sequences were obtained when the library of another leech, *Helobdella robusta*, was queried with the newly obtained *Hirudo* sequence. In addition, we did not detect homologous sequences in libraries of the cephalochordate *Branchiostoma floridae*, the tunicate *Ciona intestinalis*, the giant owl limpet *Lottia gigantea*, the starlet sea anemone *Nematostella vectensis*, the pufferfish *Takifugu rubripes*, the placozoan *Trichoplax adhaerens*, the annelid *Capitella teleta*, and the sea slug *Aplysia* or the blood fluke *Schistosoma*, the last three with whom *Hirudo* shares many similarities in protein sequences (Macagno et al. 2010).

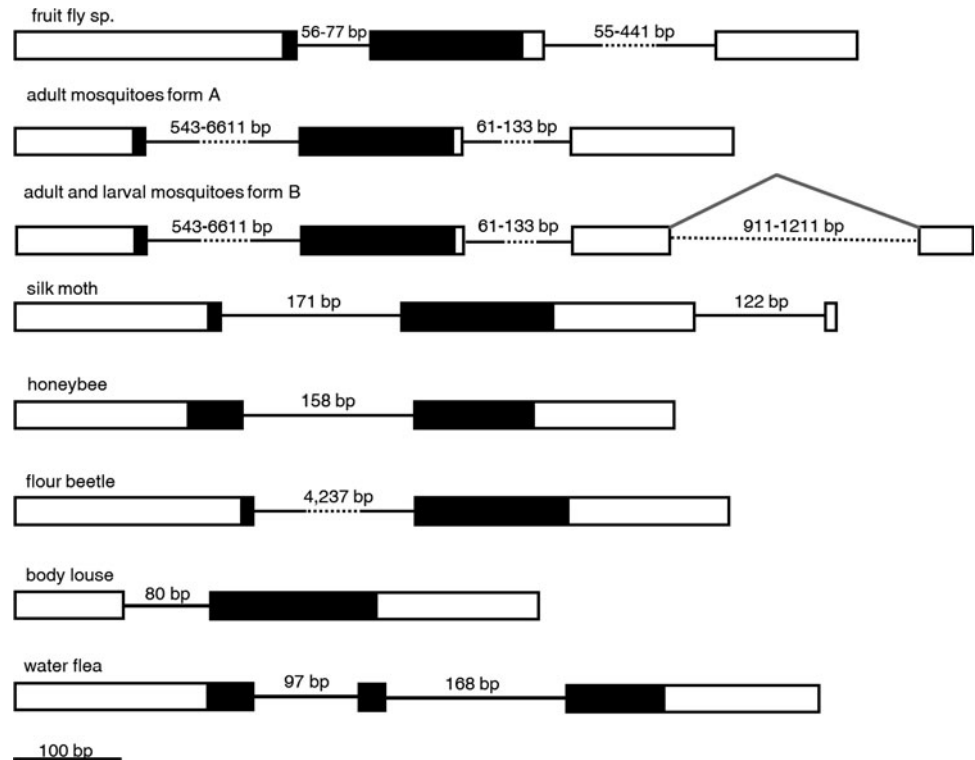
The retrieved *Hirudo* sequence (FP644758) originated from an adult CNS EST library (Macagno et al. 2010). The sequence has 100 % identity with a cDNA sequence from the sandfly *Lutzomyia longipalpis* (AM099692) that we had previously identified as flightin (Supplementary Fig. 2). In addition, when the *Hirudo* sequence is used as a query for a BLAST search, the most significant hits are from insect species and no hits were obtained from annelids. To determine if the sequence belongs to *Hirudo*, *Lutzomyia*, or both, we isolated RNA from three *Hirudo* segments and from the CNS and conducted RT-PCR analysis using primers specific for the *Hirudo* sequence. No PCR products were obtained with the flightin primers in any of the four samples tested. A PCR product of the expected size (~700 base pairs) was obtained with primers for actin indicating that the isolated RNA is suitable for reverse transcription (Supplementary Fig. 3). When the *Hirudo* primers were used to amplify cDNA obtained from the sand fly, two PCR products of approximately 500 bp were obtained (Supplementary Fig. 3). DNA sequence analysis of one of these products revealed almost 100 % identity to the *Lutzomyia* reference sequence AM099692. We conclude that sequence FP644758 did not originate from *Hirudo* and that flightin is not present in annelids.

The conserved sequences from species other than *Drosophila* sp. were used as queries to determine if additional flightin-related sequences were found outside of Arthropoda. Specifically, libraries of representatives of phyla Chordata, Mollusca, Cnidaria, and Placozoa were queried using the bristletail (*Tricholepisma*) and blue crab (*Callinectes*) sequences, but the search did not retrieve matching sequences.

### Flightin Sequence Diversity and Gene Organization

Among the species for which the complete flightin sequence is available, the protein ranges in length from 87 amino

**Fig. 1** Genomic organization of flightin coding region. Exons are indicated as *boxes* and introns as *solid lines*. Numbers indicate the length of introns in base pairs (bp). *Filled boxes* represent the conserved WYR sequence. An alternative splice variant is present in the larva of mosquitoes (*gray line*). For all sequences, the 5' end is to the left. The following sequences were used (see Supplementary Table 1): Fruit flies, all *Drosophila* species; mosquitoes, *Anopheles gambiae*; silk moth, *Bombyx mori*; honeybee, *Apis mellifera*; flour beetle, *Tribolium castaneum*; body louse, *Pediculus humanus*; water flea, *Daphnia pulex*. Scale bar 100 bp



acids in the mosquito *Culex quinquefasciatus* (form B) to 209 amino acids in the brown planthopper *Nilparvata lugens* (Supplementary Fig. 1). The differences in length are largely outside the conserved middle region, defined as the section starting at histidine 84 and ending at threonine 135. The largest sequence length differences across taxa are in the amino terminus, upstream of the conserved middle region. There are 31 different length N-termini ranging from 30 amino acids in the springtail *Onychiurus arcticus* to 89 amino acids in the fly *Drosophila willistoni*. In contrast, there are 16 different length C-termini with the shortest form of flightin in the mosquito *Culex quinquefasciatus* (three amino acids) and the longest form in the brown planthopper *N. lugens* (90 amino acids). Average length is similar in both termini, 58 amino acids for N-terminus sequences and 51 amino acids for C-terminus.

A prominent feature of N-terminus sequences is the presence of an acidic residue-rich region (amino acids 3 through 17), followed by an alanine and proline-rich region (amino acids 18 through 70) (Supplementary Fig. 1). In species with the shortest N-termini (e.g., mosquitoes), the acidic residue-rich region is reduced (e.g., mosquitoes in genera *Anopheles* and *Aedes*) or absent (e.g., *Culex quinquefasciatus*), as is the alanine–proline rich region. In contrast, some hemipterans (sharpshooters *Homalodisca*, *Oncometopia*, and *Graphocephala*) lack the acidic residue-rich and extended proline–alanine rich regions. Species with long N-termini (e.g., *Drosophila* spp.) tend to have a

high content of acidic amino acids. By contrast, the C-terminus does not exhibit any clear pattern of amino acid composition.

The coding region of the flightin gene for all *Drosophila* species, silk moth, and water flea (*Daphnia*) spans three exons separated by two introns (Fig. 1). The coding region in honeybee, flour beetle and body louse spans two exons separated by a single intron. The malaria and dengue mosquitoes have three or four coding exons, depending on the isoform expressed (see below). In most species the conserved region is encoded by two exons, but in *Daphnia* it is encoded by three exons and in the body louse it is encoded by a single exon (Fig. 1). In all species examined the conserved region is located near the middle of the open reading frame, similar to *Drosophila* flightin. However, the position of the intron that interrupts the conserved coding region is variable, as is the length.

#### Identification of a Conserved Motif in Flightin

Figure 2 shows the alignment for the most highly conserved region in flightin. This region is represented in a larger number of species in the EST database than the full-length open reading frame, hence the alignment in Fig. 2 includes more species than the alignment in Supplementary Fig. 1. In all, but one hexapod the conserved region is 52 amino acids long (in the Protura *Acerentomon franzi* it is 53 amino acids), but in crustaceans it varies in length from





	8	9	9	1	1	1	1	1	1	1	1
	4	0	5	0	5	0	5	0	5	0	5
Conserved Sites	W	R	Y	y	Y	g	P	e	R		
Blatella_germanica	.S..T-R-...	R.L.D.H.....	L.R.T-----	K.IS-.D..A.....	M... [52]						
Reticulitermes_flavipes	.S..T-S-F..D.N.D.R...	N.....	L.R.V-----	K.LN-.D..A.....	[52]						
Coptotermes_formosanus	.S..T-H-V..D.I.D.R.....	L.R.V-----	K.LN-.D..T.....	[52]							
Locusta_migratoria	Y.S..T-N-....	L.D.H.....	S.....	L.NRS-----	R.IR-SD.....	A... [52]					
Tricholepisma_aurea	.K..T-....	G.L.D.H.....	N..N.L.QRS-----	R.YK-PD.....	S... [52]						
Lepismachilis_ysignata_6	.K..T-....	ISDF.Q...S.L...L.SRS-----	R..R-QDA.K.Y...A..Y [52]								
Lepismachilis_ysignata_7	.K..T-....	ISDF.Q...S.L...L.SRS-----	R..R-QDG.K.Y...A..Y [52]								
Lepismachilis_ysignata_8	.K..T-....	ISDF.Q...S.L...L.SRS-----	R..R-QDA.Q.Y...A..Y [52]								
Onychiurus_arcticus	.Q...-S-QL.E.N.D.GS...	RPM...L.ERA-----	Q.KR-P.A.K.LF.E..A.K.S [52]								
Acerentomon_franzi	.D..R-S-AI.R.N.D.S.S..RP.L..L.GVH-----	S.RKNMA...SF...A.TS [53]									
Campodea_fragilis	.N...-S-HL.E.N.D.GSY..QPQ.K.LNQRS-----	M..K-L.T..AL..G..A.NS [52]									
Daphnia_pulex	Y.K...-S-HL.E.N.D.GS...KGM...L.ERSS-----	CK-PAP.KALN...A.K. [52]									
Daphnia_magna	.K...-S-HL.E.N.D.GS...KGM...L.ERSS-----	CK-PSP.KALN...A.K. [52]									
Daphnia_carinata	.K...-S-HL.E.N.D.GS...KGM...L.ERSS-----	CK-PSP.KALN...A.K. [52]									
Artemia_franciscana	.Q...-S-QI.E.NNDF.Q...SH.V...RR-----	D.YK-P...S...A... [52]									
Triops_cancriformis	R...-S-HV.E.N.D.GL...KNMV..V.QRNR-----	YQ-P.V.K.L...A.N. [52]									
Eurydice_pulchra	.N.R.-P-KF.D.NFD.GER..SPIVN...G.N-----	L.AH-DD...RKAF...GIDL [52]									
Gammarus_pulex_A	.D.R.-A-RF.D.NFD.G...SSMVG.V.Q.NT-----	S.--TV..RMAF...GMYS [51]									
Gammarus_pulex_B	.D.R.-A-RF.D.NFD.G...SSMVG.V.Q.NTVFSTQ.S.--TV..RMAF...GMYS [56]										
Litopenaeus_vannamei	.N.R.-A-RF.D.NWD.GD...SSMVK.M.TRS-----	DR..RMAF...G.I.S [48]									
Penaeus_monodon	.N.R.-A-RF.D.NWD.GD...SSMVK.M.TRS-----	DR..RMAF...G.I.S [48]									
Fenneropenaeus_chinensis	.N.R.-C-RF.D.NWD.GD...SSMVK.M.TRS-----	DR..RMAF...G.I.S [48]									
Homarus_americanus	I.N.R.-P-RF.D.NWD.GD...SSLVK.M.SRA-----	DR..RMAF...G.I.S [48]									
Petrolisthes_cinctipes	.N.R.-P-RF.D.NWD.GD...SSLVK.M.SR-G-----	DV..RMAF...G.I.S [48]									
Carcinus_maenas	.N.R.-P-RF.D.NWD.GD...SNLVK.M.DRSGIM--R.SQFSDM..RMAF...G.I.S [56]										
Callinectes_sapidus	.N.R.-P-RF.D.NWD.GD...SNLVK.M.DRSGIM--R.PSVPDM..RMAF...G.I.S [56]										
Carcinus_maenas_CpW	K.-.R.LPYRI.D.N.KVGES..NPQTR..EERDN----RCMK-SDP.ECASF...FAAR [54]										
Petrolisthes_cinctipes_CpW	K.K.-.LPCRI.D.N.KVGES..NPQTG..ESRDI---QRN-RV-VP.ECASY...FATK [54]										
Homarus_americanus_CpW	K.K.-.LPYRI.D.N.RVGES..NPQT...ENRELL---R-SK-QQP.EAASY...FANK [54]										
Marsupenaeus_japonicus_CpW	K.K.-.VPTRI.D.N.KVGES..NPQTEF.ENREFL-TTR--K-APP.EA..Y...FASR [55]										
Loxosceles_laeta_QpW	S.R...-YTSTI.NEN.Q.G.KL.SHTLSD.E.RYN-----	DSLS.TQV.S-DRPDLT.S. [52]									
Aphonopelma_sp_QpW	T.R...-YTSTI.NEN.R.G.GL.S.TLSD.E.RYA-----	DSLS.TQL.SD-RGDLAFS. [52]									
Tetranychus_urticae_QpW3	A.K...-YTSTI.NDN.R.GNSL.SSAL.D.ER.YN-----	ESMAKTSY.S-DRPDLG.SS [52]									
Tetranychus_urticae_QpW8	A.K...-YTSTI.NDN.R.GNSL.SSAL.D.ER.YD-----	ESMAKTSY.S-DRPDLG.SS [52]									
Psoroptes_ovis_QpW	V.K...-HTSI.NDN.R.GNSL.S.AV.E.ER.YN-----	DALS.TSF.R-DRPDLGFSS [52]									
Limulus_polyphemus_QpW	F.R...-LT.M.NHN..G.GL.SSAL.D...RYG-----	NSLS.IPLDGGPIM.LG.NL [53]									
Ixodes_scapularis_QpW	N.K...-LT.I.NDN.K.GNSL.SGA..D.E.RYQ-----	ASMS.TAF.SD-RGDLG.T. [52]									
Conserved sites	W	R	Y	y	Y	g	P	e	R		

Fig. 2 continued

48 to 56 amino acids. All indels in Crustacea can be allocated to a region spanning amino acids 116–120.

Comparison of the conserved region across all species shows there are at least six, but possibly as many as nine invariable amino acids (Fig. 2). The conserved region is delimited by an invariant tryptophan (W) at position 85 and an arginine (R) at position 131 and includes two closely spaced and invariant tyrosines (Y) at positions 93 and 104. We refer to this conserved region as the WYR motif. Two additional sites, R87 and P123, are also invariant while two other sites, Y103 and E130, are strictly conserved except for one species represented in Genbank by a single EST each. Y103 is replaced by C103 in the moth *Agrotis segetum*, but the replacement can be accounted for by a single transition in the second codon position (TGC–TAC). Likewise, E130 is replaced by Q130 in the weevil *Dia-prepes abbreviatus*, but the change results from a transversion in first codon position from GAA to CAA. These particular replacements in *A. segetum* and *D. abbreviatus* may represent sequencing errors. Position Y99 shows a single replacement to F, in the microcoryphian

*Lepismachilis y-signata* and the brine shrimp *Artemia franciscana*. In the brine shrimp the replacement also results from a single transversion (A–T) in the second codon position, but in this instance all seven (presumably independent) clones deposited in Genbank share the same sequence. Position G117 may be functionally conserved, but the alignment requires a subjective placement of gaps in decapods and isopods to bring it in line with hexapods and other crustaceans. The level of conservation in WYR is more extensive than suggested by the invariant sites alone. Four positions (108, 111, 114, and 133) show conserved replacements across all species, and two others (97 and 130) show single non-conserved replacements. In all, 23 % of the sites are conserved across all species. Among insects, 48 % of WYR sites are conserved: 15 sites are invariable and 10 show conserved replacements.

None of the amino acids that constitute WYR are emblematic of species with asynchronous flight muscle. All identities and all conserved replacements in species with asynchronous flight muscle are also identities and conserved replacements in three moth species. However, some

replacements represent unreversed apomorphies for particular clades. Position 106 (aspartic acid) is unique to insects, and positions 88 (proline) and 128 (tryptophan) are uniquely shared by hexapods and Branchiopods (*Artemia*, *Triops*, and *Daphnia*). In all other crustaceans, position 88 is a conserved arginine and position 128 is a conserved phenylalanine. Position 105 has experienced non-conserved replacements in all species outside the higher insects, and position 134 shows a non-conserved replacement in five species (*A. franzi*, dipluran *Campodea fragilis*, tadpole shrimp *Triops cancriformis*, amphipod *Eurydice pulchra*, and amphipod *Gammaru pulex*). Four positions are variable conserved replacements: 108 (non-polar), 111 (non-polar), 114 (basic, except *A. franzi*), and 133 (non-polar, except in beetle *Diabrotica virgifera*).

Other sites show different patterns of conservation. Position 84 is histidine in all species except in *Locusta migratoria* (tyrosine), *Daphnia pulex* (tyrosine), *T. cancriformis* (arginine), and lobster *Homarus americanus* (isoleucine). Position 91 is leucine in most insects and is basic (histidine, arginine, or lysine) in all crustaceans except *A. franciscana* (where it is glutamine, as in the springtail *O. arcticus*). Position 95 is a highly conserved tyrosine except in the aphids (*Toxoptera citricida*, *Myzus persicae*, *Aphis gossypii*, and *Acyrtosiphum pisum*) where it is replaced by isoleucine. Position 96 is a conserved methionine in all dipterans and *N. lugens*, an isoleucine or leucine in all other insects (except the termite *Reticulitermes flavipes*, an asparagine), and an asparagine in springtails, proturans, diplurans, and all crustaceans. Lastly, position 89 is non-polar (leucine) only in Hymenoptera.

#### Evidence for Multiple Flightin Isoforms

Mosquitoes (*A. aegypti*, *A. gambiae*, and *C. quinquefasciatus*), sandfly (*Phlebotomus papatasi*), and amphipod (*Gammarus*) are represented by two types of ESTs. In mosquitoes and the sandfly the two isoforms differ in the C-terminus, downstream from WYR; in mosquitoes the isoforms differ by alternative splicing of the distal part of the C-terminus beginning at amino acids 3, 32, and 35 downstream of WYR for *C. quinquefasciatus*, *A. gambiae* and *A. aegypti*, respectively, whereas in the sandfly the alternative splicing begins at the third amino acid downstream from WYR (Supplementary Fig. 1). The longer isoform (A in Supplementary Fig. 1) of flightin in mosquitoes is more similar to flightin in most other insects, while the shorter isoforms (B) are mosquito-specific. The mosquito *Armigeres subalatus* is represented in Genbank by a single EST of the mosquito-specific shorter isoform. Unlike mosquitoes, the shorter isoform in sandflies is more similar to flightin in other flies. Two alleles of *C. quinquefasciatus* WYR differ only at position 120 (alanine or

glycine) and three alleles of *L. y-signata* differ at positions 122 (alanine or glycine) and 124 (lysine or glutamine) (Fig. 2).

The amphipod is unique in that isoforms differ in the length of the WYR motif (indicated as A and B in Fig. 2 and Supplementary Fig. 1). In the shorter isoform (isoform A, 139 amino acids) WYR is 51 amino acids long, one less than in insects and *Daphnia*, while in the longer isoform (isoform B, 154 amino acids) WYR is 56 amino acids long, the same length as the sequence in crabs.

#### Flightin is Expressed in Larva of Some Insects

Unlike *Drosophila* where flightin expression is exclusive to the pupal and adult IFM, some ESTs from *Locusta*, silk moth *Bombyx*, *Aedes*, and dung beetle *Onthophagus* were obtained from whole nymphs or larvae, indicating that in these species flightin expression begins earlier in development. In larval *Aedes*, only the mosquito-specific B isoform of flightin is available in Genbank. In *Locusta*, *Bombyx*, and *Onthophagus* the only flightin ESTs available are those obtained from larval instars and it is not possible to determine if different instars express different isoforms. In all other instances the information associated with EST sequences in Genbank is not sufficiently specific to be useful in determining tissue or instar-specific patterns of flightin expression. Most sequences were obtained from mRNA extracted from a pool of whole larval and adult tissues or only from adults.

#### Paraflightin

Alignment of amino acid sequences revealed the presence of a second flightin-related gene in several decapod species. We designated this gene as paraflightin because it has a longer open reading frame (at least 235 amino acids) and shares similarities with flightin only in the WYR region (indicated as CpW in Fig. 2). The six invariant sites in flightin WYR (W85, R87, Y93, Y104, P123, and R131) are also invariant in paraflightin (Fig. 2). Of the positions that show a single replacement in flightin WYR, Y99 is not conserved in paraflightin WYR (paraWYR) whereas Y103 and E130 are conserved. G117 is not conserved in paraflightin. Other positions that show conserved replacements in flightin WYR (I111, K114, and L133) also experience conserved replacements in paraWYR. Among species divergence in paraWYR is only 28 %, whereas among species divergence in WYR reaches 44 %. However, this difference could simply be due to the much larger number of WYR sequences. The average divergence between WYR and paraWYR is 70 % (54–79 %) of aligned positions. Flightin and paraflightin also differ in the position of the WYR domain with respect to the full length open

reading frame; in paraflightin the conserved motif is located close to the COOH terminus while in flightin the conserved motif is always located near the middle of the open reading frame.

We conducted additional BLAST searches of the EST library using a consensus of the complete open reading frame of the paraflightin ESTs from the green crab *Carcinus maenas* and found matching sequences in the crabs *Callinectes sapidus* and *Gecarcoidea natalis*. These additional ESTs do not include the paraWYR motif and are not included in the analyses. The longest consensus of paraflightin ESTs are 263 and 235 amino acids long in *G. natalis* and *C. maenas*, respectively, but the actual length of the gene remains unknown because neither EST consensus includes a 3' stop codon. Searches of the non-redundant database using sequences of *C. maenas* and *G. natalis* did not find any significant matches.

Additional paraWYR-like sequences were identified in six species of chelicerates (QpW, Fig. 2). Only four of the six invariants sites are conserved in chelicerate paraWYR (W85, R87, Y93, and Y104). Unlike decapod paraWYR, Y99 is conserved in the chelicerates. The chelicerate ESTs differ in length and none appear to be complete open reading frames. The position of paraWYR in chelicerates tends to be closer to the N-terminus of the recovered sequences.

#### Phylogenetic Analysis

A consensus Bayesian tree estimated based on amino acid sequences of WYR and paraWYR (CpW) using a mixed model of evolution (as implemented in Mr. Bayes 3.2, (Ronquist et al. 2012) is shown in Fig. 3. The sequences of the WYR-like domain present in chelicerates were not included in the analyses shown. When sequences found in chelicerates were considered, the Bayesian analyses did not converge and results of the four runs produced largely incongruent or otherwise unresolved consensus trees. Phylogenetic analysis of DNA sequences including chelicerates produced almost completely unresolved trees, most likely due to rampant saturation over the short sequences (results not shown).

All Bayesian analyses support the monophyly of families Drosophilidae (represented by 15 species), Aphidiidae (4 genera), and Cicadellidae (3 genera), the orders Hymenoptera (14 genera), Lepidoptera (3 species), and Dictyoptera (roaches and termites, 3 genera), the class Insecta, and the sequences representing paraWYR. Some well-supported relationships are the placement of Dictyoptera in the clade containing the Paraneoptera (lice, aphids, leafhoppers, and true bugs) and Holometabola orders, the sister group relationship between the green lacewing *Chrysoperla* and Hymenoptera, the sister group relationship

**Fig. 3** Bayesian consensus tree based on analysis of the amino acid sequence of WYR and paraWYR (CpW). Numbers above the branches refer to posterior probabilities resulting from four runs of 20 million generations and 25 % burnin. Dipterans are indicated in *bold* and decapod paraWYR in *italics*. Insect orders and other taxa retained as monophyletic are identified by *capital letters*

between *Artemia* and insects, and the placement of Collembola, Protura, and Diplura as sister to the clade comprising insects, *Artemia* and *Triops*. Support for other taxa is weak (less than 90 posterior probability) or absent.

Lack of support for some groups is also noticeable. There is no structure in the clade containing the higher insects, whereas the orders Hemiptera, Coleoptera, and Diptera are never monophyletic, and even mosquitoes fail to form a monophyletic group. Similarly, although not unexpectedly, the crustaceans do not form a monophyletic group, and even the malacostracan clade (decapods, isopod, and amphipod) shows non-significant support.

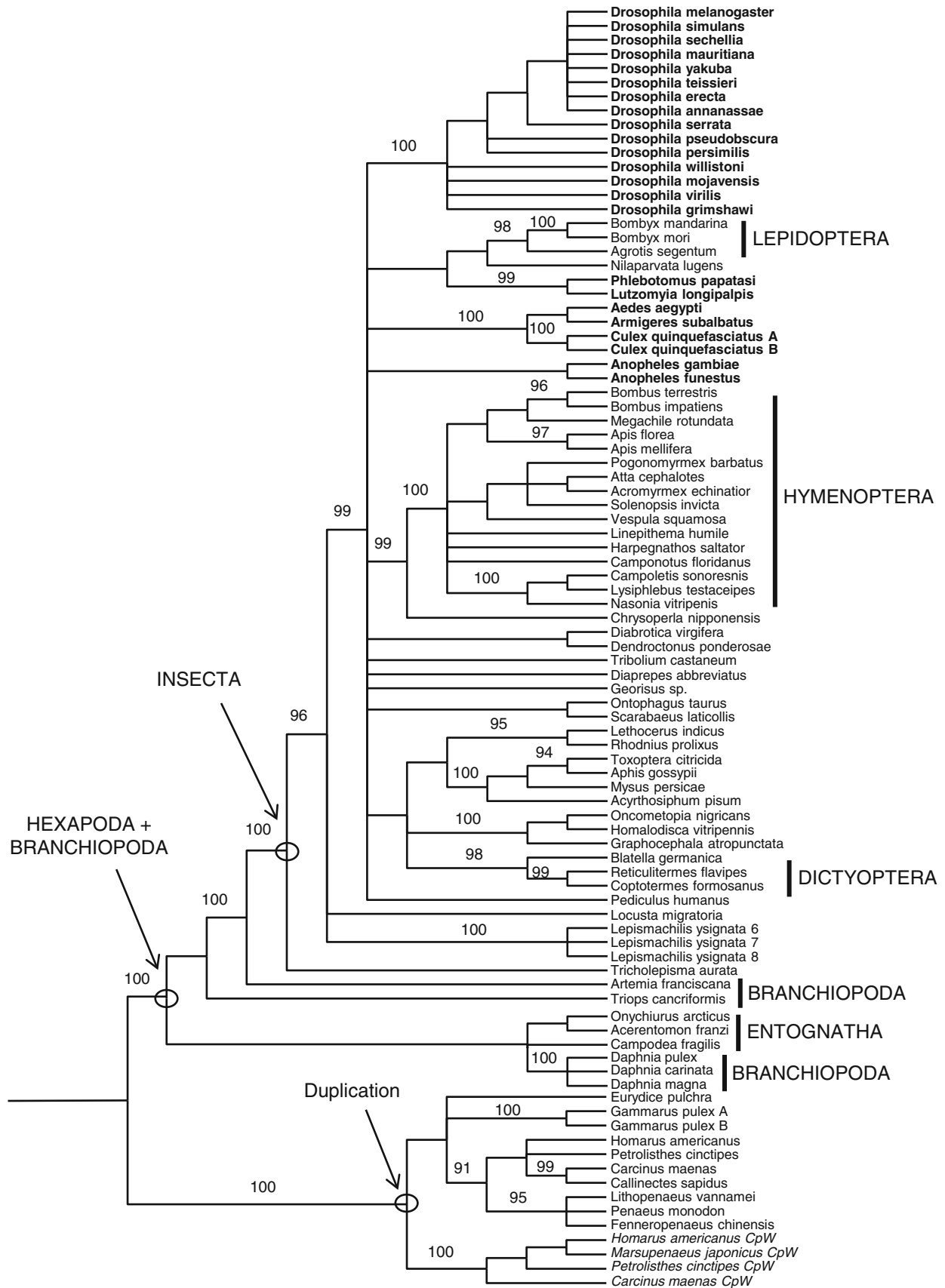
#### Discussion

This study represents the most complete phylogeny of flightin to date. Previous work in *Drosophila* (Vigoreaux et al. 1993) and *Lethocerus* (Ferguson et al. 1994) had shown that flightin expression is limited to IFM. Here, we identified flightin sequences in crustaceans, wingless hexapods, and larva of winged insects, indicating that flightin is a protein with deep ancestry and functions outside of flight muscles.

#### Evolution of Flightin

The conservation of the WYR sequence together with the similarity in gene structure, length of the open reading frame, and relative position of the WYR region within the open reading frame provide evidence that the genes identified in this study are flightin. In most species the WYR domain is encoded by two exons, the two exceptions being the water flea (three exons) and the body louse (one exon). The position of the intron within the WYR sequence is not conserved suggesting that insertion of introns occurred independently in most genes. The conservation of the WYR sequence stands in sharp contrast to flightin sequences outside this domain, which show a faster rate of evolution and cannot be reliably aligned between insect orders. In *D. melanogaster*, the sequences C-terminal to WYR have been shown to be required for normal flightin function. Specifically, transgenic *Drosophila* expressing a truncated flightin with a C-terminus that extends only 4 amino acids beyond the WYR region (i.e., missing the last 44 amino acids, *fln*<sup>ΔC44</sup> (Tanner et al. 2011)) are flightless as a result of decreases in myofibril lattice organization





and power output. In contrast, mosquitoes in general, and *C. quinquefasciatus* in particular, express flightin with short C-terminus sequences. The adult B isoform of the southern house mosquito has a three amino acid C-terminus region. Thus, the C-terminus region is likely involved in taxon-specific roles that respond to different evolutionary selective forces than the WYR region. A similar argument could be made for the N-terminus region which shows even more variability in length and amino acid sequence than the C-terminus region.

The presence of flightin in branchiopods, proturans, dipurans, springtails, and bristletails indicate that this protein originated over 500 million years ago. This taxonomic distribution, together with the recovery of ESTs from larval tissues, indicates flightin has functions beyond its well-characterized role in insect IFM. There is evidence that insects date back to approximately 400 million years ago, although flight itself most likely arose during the late Devonian, ~360 million years ago (Grimaldi and Engel 2005). This raises questions as to when, how, and why flightin evolved into an IFM-specific protein in some taxa (*Drosophila* and *Lethocerus*), but not in most others. The fact that flightin is not IFM-specific in at least some dipterans, as evidenced by its presence in mosquito larva, suggest that IFM specificity is a relatively recent event that may have occurred independently in fruit flies and waterbug.

The origin of the Pancrustacea clade has been estimated to date to the Precambrian, >600 million years ago (Regier et al. 2005). The Cambrian period is well known for its “explosion” of new phyla, a period during which an enormous diversity in body plans was prevalent, including the ones that gave rise to Arthropoda. While many of the Cambrian phyla have disappeared, Arthropoda and Pancrustacea in particular, today continue to be defined by diversity in body plan and additionally, by changes in body structure during development (molting and metamorphosis). This flexibility in body plan is one of the major contributors to the hyper-diversification of hexapods and crustaceans given their ability to exploit many different environments as crawlers, walkers, jumpers, flyers, or swimmers. The various body types and modes of locomotion are by and large supported by muscles that have evolved specialized functional attributes. The results presented here suggests flightin may have facilitated the diversification of hexapods and crustaceans by serving as an adaptable component of the muscle contractile structure that is otherwise characterized by the highly conserved nature of its major components including actin, myosin, and tropomyosin among others (Sheterline et al. 1998; Kreis and Vale 1999; Ruiz-Trillo et al. 2002; Ayme-Southgate et al. 2008; Barua et al. 2011). This view is consistent with that of most taxonomically restricted genes, also referred to as orphans, that have been proposed to

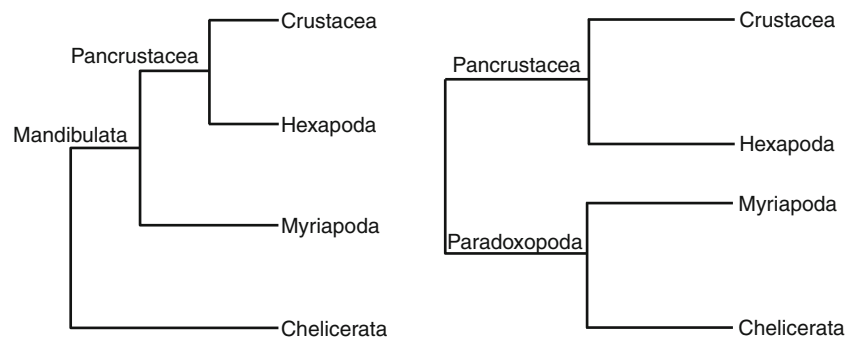
function in lineage-specific adaptations (Domazet-Loso and Tautz 2003).

### The WYR Region Represents a New Protein Domain

In *D. melanogaster*, the function of flightin is dependent on its interaction with myosin. The *Mhc*<sup>13</sup> mutation in the light meromyosin (coiled coil) region of the myosin rod prevents IFM accumulation of flightin in vivo and myosin binding in vitro (Kronert et al. 1995; Ayer and Vigoreaux 2003). While the precise myosin binding site in flightin has not been identified, results from our laboratory suggest the binding site resides within or overlaps the WYR domain. Neither the aforementioned *fln*<sup>AC44</sup> (Tanner et al. 2011) nor a deletion of 62 amino acids N-terminal to the WYR region (Chakravorty 2013) affect flightin accumulation in the IFM, suggesting that both N-terminus and C-terminus truncated flightins are stable proteins and are incorporated into the thick filament. Thus, we propose that the WYR motif defines a new class of myosin coiled coil binding domain.

The alpha helical coiled coil is among the most common protein motifs involved in a wide range of protein interactions (Wang et al. 2012). Despite its ubiquitous nature, very few coiled-coil binding motifs are known. Two motifs have been identified that bind the light meromyosin region of muscle myosin II. Vertebrate myosin binding protein C (MyBP-C) binds to the myosin rod through its C-terminus immunoglobulin (Ig) domain, referred to as CX (Miyamoto et al. 1999). In vitro, the CX domain binds to a region of the myosin rod that overlaps with, or is identical to the myosin region that binds flightin (Ayer and Vigoreaux 2003; Flashman et al. 2007). The region of the myosin rod that binds flightin and MyBP-C is contained within the region that binds My1, a unique N-terminus domain in myomesin, an elastic protein of the vertebrate M-line (Obermann et al. 1997). Thus, it appears that flightin, MyBP-C, and myomesin have evolved distinct domains to fulfill a common function of binding to the coiled coil. The comparison between flightin and MyBP-C should prove particularly interesting because these two proteins are known to augment the stiffness of the thick filament in insect flight muscle and vertebrate cardiac muscle, respectively (Nyland et al. 2009; Contompasis et al. 2010). This raises the possibility that WYR and CX are analogous functional motifs. Given the impact of thick filament mechanical properties on muscle performance (Miller et al. 2010), the interaction of flightin and MyBP-C with the myosin coiled coil can be seen as convergent evolutionary strategies to influence muscle functional output. However, preliminary structure predictions of the WYR motif predict a protein fold different from the immunoglobulin domain fold indicating that convergence is not through a common fold and that WYR represents a novel protein domain.

**Fig. 4** Phylogenetic hypotheses of arthropod relationships (modified from Giribet and Edgecombe 2012)



### Implications for Arthropod Phylogeny

There are competing hypotheses for the phylogeny of arthropods, especially with respect to the relationship among the major sub-phyyla (Giribet and Edgecombe 2012) (Fig. 4). Recent phylogenetic analysis of 62 nuclear protein-coding genes places Chelicerata as sister to a reorganized Mandibulata, in which Myriapoda is sister to Pancrustacea (Tetracnata) (Regier et al. 2010). In contrast, Paradoxopoda (Myriochelata) places Myriapoda and Chelicerata as one clade sister to Pancrustacea (Mallatt et al. 2004; Dunn et al. 2008). Flightin (WYR) is present in hexapods and crustaceans and absent in myriapods and chelicerates, though this latter group possesses a sequence that resembles paraWYR. The distribution of flightin thus supports Pancrustacea and is not informative with respect to Mandibulata. The paraphyly of Hexapoda (Nardi et al. 2003) and reciprocal paraphyletic relationship of Hexapoda and Crustacea (Carapelli et al. 2007) are also reflected in the flightin phylogeny. In particular, the sister group relationship of Hexapoda and Branchiopoda, supported by a number of molecular studies (reviewed in Grimaldi 2010), is also evident in the WYR phylogeny. One caveat of this study is that taxa are not uniformly represented in the database; the number of myriapod species (~34) and sequences (4,466 EST, 9,548 other DNA sequences) is far smaller compared to insects and to a lesser extent, crustaceans.

The phylogeny generated from WYR sequences lacks the resolution to make definitive statements about evolutionary relationships. While some of the relationships recovered are consistent with those previously recognized (e.g., the monophyly of Insecta), there is widespread lack of structure in the insect clade, including paraphyletic Holometabola (orders Diptera, Coleoptera, Lepidoptera, and Neuroptera) and Diptera. This failure to recover strongly supported monophyletic groups may result from the limited number of amino acids (~50) in the sequence. It is possible that the few positions allowed to vary have experienced saturation and for purposes of the analysis they are randomized. However, one would expect that lack of signal due to short and fast evolving sequence would

also affect the monophyly of Hymenoptera and even more deeper nodes such as the monophyly of the insects, the sister clade relationship between *Artemia* and insects, and the clade that includes Branchiopoda (*Artemia*, *Triops*, and *Daphnia*) and Hexapoda. Yet, these clades, previously supported by molecular and morphological studies (reviewed in Jenner 2010), are also supported by the current analysis. Thus, another interpretation of our results is that flightin is evolving at a different rate than the genome in general, an interpretation that is further supported by the poor conservation of sequences outside the WYR region.

An alternative explanation for the absence of resolution in the insect clade is that WYR is under different levels of selective pressure in different groups. The hymenopteran genera sample represents very different life histories and strategies, from small parasitoids (e.g., *Lysiphlebus testaceipes*) to large bees (e.g., *Bombus terrestris*) to ants (e.g., *Harpegnathos saltator*), but despite these differences all are included in a strongly supported clade. It is possible that in this group selective pressures regulating nucleotide composition have resulted in well-conserved amino acid evolution. The lack of support for the monophyly of beetles, bugs, Diptera, and even mosquitoes may reflect either relaxed or disruptive evolution of WYR, leading different lineages within each order to different optima.

The choice of root placement on the consensus tree in Fig. 3 suggests that the duplication of the ancestral gene that gave rise to paraWYR and WYR took place after the differentiation of Malacostraca. Elucidation of the actual pattern of relationships between paraWYR, WYR, and the WYR-like domain in chelicerates may have to await the identification of an appropriate outgroup. Further insight into this issue and the final elucidation of the origin of WYR family of domains will require more sophisticated functional sequence analysis of a greater representation of genes across a larger set of taxa.

In summary, the flightin gene consists of a relatively well-conserved domain flanked by two faster evolving domains. Flightin is a distinct feature of Pancrustacea present in a single copy per genome, and only in Decapoda is there evidence of a paralogue. The origin of flightin

precedes the split of hexapods from crustacean ancestors, estimated at 420–480 million years ago (Grimaldi 2010). While the function of flightin outside *Drosophila* remains to be determined, our results indicate that flightin did not originate for an IFM-specific function in higher order insects. Flightin may prove to be effective for investigating the evolution of protein specialization in the highly diverse and speciose pancrustacean evolutionary lineage.

## Materials and Methods

### Sequence Retrieval and Alignments

The *D. melanogaster* flightin sequence [(Vigoreaux et al. 1993), accession # AY060802] was used to search the genomes of eleven additional species of *Drosophila*, honeybee (*Apis mellifera*), parasitoid wasp (*Nasonia vitripennis*), dengue and malaria mosquitoes (*Aedes aegypti* and *Anopheles gambiae*, respectively), flour beetle (*Tribolium castaneum*), silk moth (*Bombyx mori*), body louse (*Pediculus humanus*), and water flea (*Daphnia pulex*). Flightin sequences for *Drosophila mauritiana* and *Drosophila teissieri* were generated via PCR amplification (see below). Fly stocks were obtained from *Drosophila* Species Stock Center (formerly at the University of Arizona, now at the University of California San Diego).

Additional examples of flightin were obtained from GenBank. All GenBank BLAST searches were initially performed using each Blosom matrix (45, 62, 80), without taxonomic restrictions and without enforcing the low complexity filter. First, we searched the non-redundant amino acid database (i.e., all non-redundant GenBank CDS translations, RefSeq proteins, PDB, SwissProt, PIR, and PRF), followed by a tblastn of the EST library (i.e., database of GenBank, EMBL, and DDBJ sequences from EST division) and the table of genomes. In addition to taxon independent searches for EST libraries, we performed taxon-specific searches for each order of Hexapoda and Crustacea, and for Classes Arachnida, Diplopoda, Chilopoda, Pauropoda, and Symphyla. Genbank was last searched 11 May 2012.

Several species in GenBank are represented by many ESTs (e.g., 83 for *A. gambiae*), some of which differ in sequence. Differences among sequences attributed to single species are either polymorphisms or sequencing errors. In order to identify potential sequencing errors, all ESTs DNA sequences available for a given species were aligned using the default parameters in Clustal W (Thompson et al. 1994). Some errors were easy to recognize because they produce ESTs with ragged ends, obvious frame shifts or premature stop codons. Single site sequencing errors are indistinguishable from true polymorphisms. Substitutions representing obvious

sequencing errors were scored following a 75 % majority rule consensus sequence constructed using MacClade 4.05 (Maddison and Maddison 2002). Sequences with substitutions that could not be dismissed as sequencing errors were included in the analyses as independent operational taxonomic units (OTU's) only if they code for different amino acids. For species represented only by ESTs, the amino acid sequence used in all analyses is based on translation of the consensus cDNA sequence.

To align amino acid sequences between species, DNA sequences were first conceptually translated using DNA Strider 1.2 (Marck 1988; Douglas 1995), and the resulting amino acid sequences were aligned using Clustal W.

### PCR and Sequencing

Flightin sequences for *D. mauritiana* and *D. teissieri* were generated from genomic DNA using PCR primers 25FD1 (5' GCR GAC GAA GAR GAY CCD TGG G 3') and 541R (5' AAG GAC ACT GGC ATA CCT TTG GTT 3'). Genomic DNA for PCR was extracted with the Puregene DNA isolation kit (Gentra Systems). PCR time and temperature profile was 3 min at 95 °C, followed by 40 cycles of 1 min at 95 °C, 1 min at 52 °C, and 1 min at 72 °C. PCR products were sequenced directly using the Big Dye sequencing kit. Sequencing temperature and time profiles were 3 min at 98 °C followed by 25 cycles of 30 s at 98 °C, 30 s at 50 °C, and 4 min at 60 °C.

Adult leeches (*Hirudo medicinalis*), purchased from Carolina Biological, were dissected in three regions: anterior, medial, and posterior body. Adult sand flies (*Lutzomyia longipalpis*, Jacobina strain), preserved in RNAlater (Qiagen), were obtained from the Laboratory of Malaria and Vector Research at NIAID-NIH. Total RNA from two adult leeches was extracted using Trizol and total RNA from 10 sand flies specimen was extracted using RNAeasy minikit (Qiagen). Isolated RNA (~3 µg) was treated with RQ1 RNase free DNase (Promega) and first strand synthesis was performed using SuperScript III reverse transcriptase (Invitrogen) following manufacturer's instructions. A sample lacking reverse transcriptase was used as control. PCR amplification was performed using primers based on the *H. medicinalis* flightin cDNA sequence (GenBank accession number FP644758):

Hir-forward: 5' ATGGCTGAGGACGATCCATG 3'  
Hir-reverse: 5' CAAAATGCTGGAATACTTCCGG 3'

The control PCR reactions were done using primers for actin:

Actin90Fw: 5' TGGCAYCAYACNNTTYTAYAA 3'  
Actin300Rv: 5' GCDATNCCNGGRTACATNGT 3'



PCR conditions were 95 °C, 3 min; 35 cycles of 95 °C, 1 min; 57 °C, 1 min; 72 °C, 1 min. PCR fragments of the expected size were gel purified and cloned into pCR4-TOPO/DH5alpha-TI *E. coli* system (Invitrogen) and the DNA sequence obtained for three independent clones.

### Phylogenetic Analysis

The Bayesian analyses were carried out with MrBayes 3.2 (Ronquist et al. 2012), using the non-redundant amino acid sequences for the conserved WYR region. The Bayesian analyses implemented a mix model of amino acid sequence evolution and ran for 20,000,000 generations and four chains. One tree was sampled every 1,000 generations, resulting in 2,001 trees per run. The Bayesian analysis was performed four times. Posterior probabilities were estimated after a 25 % burnin.

**Acknowledgments** This research was supported by National Science Foundation awards IOS-0718417 and MCB-1050834 to JOV. FNSA was supported in part by Grant HL007944 from the National Institutes of Health. PAO was supported in part by NSF 0450339 AGEF. We thank Jesus G. Valenzuela and Claudio Meneses, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, for providing *Lutzomyia*, and members of the Vigoreaux and Maughan labs for helpful discussions.

### Reference

- Ayer G, Vigoreaux JO (2003) Flightin is a myosin rod binding protein. *Cell Biochem Biophys* 38:41–54
- Ayme-Southgate AJ, Southgate RJ, Philipp RA, Sotka EE, Kramp C (2008) The myofibrillar protein, projectin, is highly conserved across insect evolution except for its PEVK domain. *J Mol Evol* 67:653–669
- Barton B, Ayer G, Heymann N, Maughan DW, Lehmann FO, Vigoreaux JO (2005) Flight muscle properties and aerodynamic performance of *Drosophila* expressing a flightin transgene. *J Exp Biol* 208:549–560
- Barua B, Pamula MC, Hitchcock-DeGregori SE (2011) Evolutionarily conserved surface residues constitute actin binding sites of tropomyosin. *Proc Natl Acad Sci USA* 108:10150–10155
- Brodsky AK (1994) *The evolution of insect flight*. Oxford University Press, Oxford
- Carapelli A, Lio P, Nardi F, van der Wath E, Frati F (2007) Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evol Biol* 7(Suppl 2):S8.1471–S8.2148. doi:10.1186/1471-2148-7-S2-S8
- Chakravorty S (2013) Role of the *Drosophila melanogaster* indirect flight muscles in flight and male courtship song: studies on flightin and myosin light chain-2. Biology. Ph.D. thesis, Burlington, University of Vermont, 339 pp
- Chan WP, Dickinson MH (1996) In vivo length oscillations of indirect flight muscles in the fruit fly *Drosophila virilis*. *J Exp Biol* 199:2767–2774
- Contompasis JL, Nyland LR, Maughan DW, Vigoreaux JO (2010) Flightin is necessary for length determination, structural integrity, and large bending stiffness of insect flight muscle thick filaments. *J Mol Biol* 395:340–348
- Dohle W (2001) Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name ‘Tetraconata’ for the monophyletic unit Crustacea+Hexapoda. *Annales de la Societe Entomologique de France* 37:85–103
- Domazet-Loso T, Tautz D (2003) An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res* 13:2213–2219
- Douglas SE (1995) DNA Strider. An inexpensive sequence analysis package for the Macintosh. *Mol Biotechnol* 3:37–45
- Dudley R (2000) *The biomechanics of insect flight*. Princeton University Press, Princeton
- Dunn CW, Hejnol A, Matus DQ et al (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749
- Ferguson C, Lakey A, Hutchings A, Butcher GW, Leonard KR, Bullard B (1994) Cytoskeletal proteins of insect muscle: location of zeelins in *Lethocerus* flight and leg muscle. *J Cell Sci* 107:1115–1129
- Flashman E, Watkins H, Redwood C (2007) Localization of the binding site of the C-terminal domain of cardiac myosin-binding protein-C on the myosin rod. *Biochem J* 401:97–102
- Giribet G, Edgecombe GD (2012) Reevaluating the arthropod tree of life. *Annu Rev Entomol* 57:167–186
- Grimaldi DA (2010) 400 million years on six legs: on the origin and early evolution of Hexapoda. *Arthropod Struct Dev* 39:191–203
- Grimaldi D, Engel MS (2005) *Evolution of the insects*. Cambridge University Press, Cambridge
- Henkin JA, Maughan DW, Vigoreaux JO (2004) Mutations that affect flightin expression in *Drosophila* alter the viscoelastic properties of flight muscle fibers. *Am J Physiol Cell Physiol* 286:C65–C72
- Hyatt CJ, Maughan DW (1994) Fourier analysis of wing beat signals: assessing the effects of genetic alterations of flight muscle structure in Diptera. *Biophys J* 67:1149–1154
- Jenner RA (2010) Higher-level crustacean phylogeny: consensus and conflicting hypotheses. *Arthropod Struct Dev* 39:143–153
- Kreis T, Vale R (eds) (1999) *Guidebook to the cytoskeletal and motor proteins*. Oxford University Press, Oxford
- Kronert WA, O’Donnell PT, Fieck A, Lawn A, Vigoreaux JO, Sparrow JC, Bernstein SI (1995) Defects in the *Drosophila* myosin rod permit sarcomere assembly but cause flight muscle degeneration. *J Mol Biol* 249:111–125
- Macagno ER, Gaasterland T, Edsall L et al (2010) Construction of a medicinal leech transcriptome database and its application to the identification of leech homologs of neural and innate immune genes. *BMC Genomics* 11:407. doi:10.1186/1471-2164-11-407
- Maddison WP, Maddison DR (2002) *MacClade: analysis of phylogeny and character evolution*, version 4.05 for Mac OS X. Sinauer, Sunderland
- Mallatt JM, Garey JR, Shultz JW (2004) Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Mol Phylogenet Evol* 31:178–191
- Marck C (1988) ‘DNA Strider’: a ‘C’ program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res* 16:1829–1836
- Maughan D, Vigoreaux J (2004) Nature’s strategy for optimizing power generation in insect flight muscle. In: Sugi H (ed) *Mysteries about the sliding filament mechanism: fifty years after its proposal*. Plenum Press, New York
- Miller MS, Tanner BCW, Nyland L, Vigoreaux JO (2010) Comparative biomechanics of thick filaments and thin filaments with functional consequences for muscle contraction. *J Biomed Biotech* 2010:14. doi:10.1155/2010/473423
- Miyamoto CA, Fischman DA, Reinach FC (1999) The interface between MyBP-C and myosin: site-directed mutagenesis of the

- CX myosin-binding domain of MyBP-C. *J Muscle Res Cell Motil* 20:703–715
- Moore JR (2006) Stretch activation: toward a molecular mechanism. In: Vigoreaux JO (ed) *Nature's versatile engine: insect flight muscle inside and out*. Springer/Landes Biosci, New York, pp 44–60
- Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: monophyletic or paraphyletic? *Science* 299:1887–1889
- Nongthomba U, Cummins M, Clark S, Vigoreaux JO, Sparrow JC (2003) Suppression of muscle hypercontraction by mutations in the myosin heavy chain gene of *Drosophila melanogaster*. *Genetics* 164:209–222
- Nyland LR, Palmer BM, Chen Z, Maughan DW, Seidman CE, Seidman JG, Kreplak L, Vigoreaux JO (2009) Cardiac myosin binding protein-C is essential for thick-filament stability and flexural rigidity. *Biophys J* 96:3273–3280
- Obermann WM, Gautel M, Weber K, Furst DO (1997) Molecular structure of the sarcomeric M band: mapping of titin and myosin binding domains in myomesin and the identification of a potential regulatory phosphorylation site in myomesin. *EMBO J* 16:211–220
- Qiu F, Brendel S, Cunha PM, Astola N, Song B, Furlong EE, Leonard KR, Bullard B (2005) Myofilin, a protein in the thick filaments of insect muscle. *J Cell Sci* 118:1527–1536
- Reedy MC, Bullard B, Vigoreaux JO (2000) Flightin is essential for thick filament assembly and sarcomere stability in *Drosophila* flight muscles. *J Cell Biol* 151:1483–1499
- Regier JC, Shultz JW, Kambic RE (2005) Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proc Biol Sci* 272:395–401
- Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463:1079–1083
- Richter S (2002) The Tetraconata concept: hexapod-crustacean relationships and the phylogeny of Crustacea. *Org Divers Evol* 2:217–237
- Ronquist F, Teslenko M, van der Mark P et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542
- Ruiz-Trillo I, Paps J, Loukota M, Ribera C, Jondelius U, Baguna J, Riutort M (2002) A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. *Proc Natl Acad Sci USA* 99:11246–11251
- Sheterline P, Clayton J, Sparrow J (1998) *Actin*. Oxford University Press, Oxford
- Tanner BC, Miller MS, Miller BM, Lekkas P, Irving TC, Maughan DW, Vigoreaux JO (2011) C-terminal truncation of flightin decreases myofilament lattice organization, cross-bridge binding, and power output in *Drosophila* indirect flight muscle. *Am J Physiol Cell Physiol* 301:C383–C391
- 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* 22:4673–4680
- Vigoreaux JO (1994) Alterations in flightin phosphorylation in *Drosophila* flight muscles are associated with myofibrillar defects engendered by actin and myosin heavy chain mutant alleles. *Biochem Genet* 32:301–314
- Vigoreaux JO, Saide JD, Valgeirsdottir K, Pardue ML (1993) Flightin, a novel myofibrillar protein of *Drosophila* stretch-activated muscles. *J Cell Biol* 121:587–598
- Vigoreaux JO, Hernandez C, Moore J, Ayer G, Maughan D (1998) A genetic deficiency that spans the flightin gene of *Drosophila melanogaster* affects the ultrastructure and function of the flight muscles. *J Exp Biol* 201:2033–2044
- Wang Y, Zhang X, Zhang H et al (2012) Coiled-coil networking shapes cell molecular machinery. *Mol Biol Cell* 23:3911–3922
- Wilson EO (1992) *The diversity of life*. W. W. Norton & Company, New York