Mitochondrial Genome Sequences of Nematocera (Lower Diptera): Evidence of Rearrangement following a Complete Genome Duplication in a Winter Crane Fly

Andrew T. Beckenbach*

Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

*Corresponding author: E-mail: beckenba@sfu.ca.

Accepted: 1 December 2011

Abstract

The complete mitochondrial DNA sequences of eight representatives of lower Diptera, suborder Nematocera, along with nearly complete sequences from two other species, are presented. These taxa represent eight families not previously represented by complete mitochondrial DNA sequences. Most of the sequences retain the ancestral dipteran mitochondrial gene arrangement, while one sequence, that of the midge *Arachnocampa flava* (family Keroplatidae), has an inversion of the *trnE* gene. The most unusual result is the extensive rearrangement of the mitochondrial genome of a winter crane fly, *Paracladura trichoptera* (family Trichocera). The pattern of rearrangement indicates that the mechanism of rearrangement involved a tandem duplication of the entire mitochondrial genome, followed by random and nonrandom loss of one copy of each gene. Another winter crane fly retains the ancestral diperan gene arrangement. A preliminary mitochondrial phylogeny of the Diptera is also presented.

Key words: mitochondrial genomics, Nematocera, dipteran phylogeny.

Introduction

The animal mitochondrial genome typically codes for 37 genes, including 13 genes for proteins involved in the electron transport system, a minimal set of 22 transfer RNAs (tRNAs) and two ribosomal RNAs (rRNAs) (Boore 1999). These genes are arranged on a very compact circular genome, arrangements that are relatively stable over long periods of evolutionary history (Boore 2000). The arrangement first encountered in the fly, *Drosophila yakuba* (Clary and Wolstenholme 1985), is now known to be widespread across insects and is likely the ancestral arrangement for the order Diptera (Boore et al. 1998; Cameron et al. 2006).

While most Diptera retain the ancestral arrangement, rearrangements are occasionally observed. Mosquitoes (family Culicidae), gall and sciarid midges (families Cecidomyiidae and Sciaridae) are known to have minor rearrangements of tRNA genes (Beard et al. 1993; Mitchell et al. 1993; Beckenbach and Joy 2009). These rearrangements include inversions, where the coding direction and strand are switched, and transpositions, where the gene is moved to another location in the genome, but the coding direction retained. Duplications of tRNA genes are occasionally observed and have been documented in blowflies (Lessinger et al. 2004). In none of the dipteran genomes previously described are there rearrangements of the major genes (those coding for proteins and rRNAs). More extensive rearrangements, involving both tRNA and major genes, have been found in other insect orders, such as thrips, order Thysanoptera (Shao and Barker 2003), and lice, order Phthiraptera (Cameron, Johnson, et al. 2007).

Diptera is one of four megadiverse orders of holometabolous insects (those that undergo complete metamorphosis). The order probably originated about 260 Ma and subsequently underwent three episodes of radiation (Wiegmann et al. 2011). The first radiation, from about 240 to 220 Ma, gave rise to an assortment of families and superfamilies collectively known as the Nematocera. The second radiation, between about 180 and 150 Ma, gave rise to the lower ("orthorrhaphous") Brachycera. The most recent radiation, between about 65 and 40 Ma, produced the "higher" Brachycera (Schizophora). The order has traditionally been divided into two suborders: Nematocera and Brachycera. It has long been understood that the Brachycera arose from within the Nematocera. Prior to this

study, complete mitochondrial genomes from only three nematoceran families have been described.

The purpose of this study was to examine mitochondrial genomes from a wide diversity of nematoceran families and superfamilies. In the course of this study, a highly rearranged genome was discovered in a species of winter crane fly (family Trichoceridae). The pattern of rearrangement provides considerable insight into the mechanisms involved in rearrangement of genes in this genome. I also use these new sequences, along with previously published sequences, to provide a preliminary mitochondrial DNA phylogeny of the Diptera.

Materials and Methods

Source Material

Adults of a false crane fly, *Ptychoptera* sp., a phantom crane fly, *Bittacomorphella fenderiana* (family Ptychopteridae), a winter crane fly, *Paracladura trichoptera* (family Trichoceridae), *Cramptonomyia spenceri* (family Pachyneuridae), and a wood gnat, *Sylvicola fenestralis* (family Anisopodidae) were collected on the campus of Simon Fraser University, Burnaby Mountain, British Columbia. Adults of the winter crane fly, *Trichocera bimaculata* (family Trichoceridae), the midges *Arachnocampa flava* (family Keroplatidae) and *Chironomus tepperi* (family Chironomidae), a larva of a crane fly, *Tipula abdominalis* (family Tipulidae), and of a primitive crane fly, *Protoplasma fitchii* (family Tanyderidae) were provided by the Dipteran Tree of Life Project.

DNA Extraction and Polymerase Chain Reaction Amplification

Legs were removed from adults of the larger species, *Ptychoptera*, *Bittacomorphella*, *Paracladura*, *Cramptonomyia*, and *Sylvicola* specimens for separate extraction. The midges, *Arachnocampa* and *Chironomus*, and the winter crane fly, *Trichocera*, were ground up as entire individuals. The *Tipula* and *Prototanyderus* larvae were cut into sections. DNA extraction was carried out using a standard phenol purification, followed by extraction with chloroform/isoamyl alcohol and ethanol precipitation (Liu and Beckenbach 1992). The pellets were washed one time with 70% ETOH and allowed to air-dry overnight. Dried samples were frozen at −20 °C until needed.

Details of the polymerase chain reaction (PCR) amplification and sequencing methods employed are given in Beckenbach (2011). Briefly, fragments between 500 and 1,500 bp were amplified using standard primers (Simon et al. 2006, Supplemental Primer List) and sequenced on both strands using the amplification primers. For fragments larger than about 800 bp, additional internal primers were chosen for further amplification and sequencing. This procedure gave partial sequence for all taxa. Additional primers were designed for each taxon to fill in the regions, which did not amplify with standard primers.

Control regions were amplified using primers SR-J14610 paired with either TM-N200 or TI-N9 (5′-TCAAGGTAAYCCTTTTTRTCAGGC), using Phusion high-fidelity DNA polymerase (Finnzymes, Finland) as described in Beckenbach (2011). Amplified products were purified and sequenced using both amplification primers. Taxon specific primers were designed as necessary to fill in gaps.

One of the winter crane fly genomes, that of *Paracladura*, is highly rearranged. The initial amplification and sequencing steps produced internal sequence for most major genes, but little information about gene organization. These sequence fragments were joined together by trial and error amplification using well-matched primers in various combinations.

Analysis

Sequences were aligned and assembled manually. Ambiguous sites were resolved by reamplifying and resequencing the region using different primer pairs and by examination of the sequencing traces. Protein coding genes were identified as open reading frames corresponding to the 13 protein coding genes expected in metazoan mitochondrial genomes. The tRNA genes were identified using tRNAscan-SE (Lowe and Eddy 1997), with a COVE cutoff score of 4. This process located 20 of the 22 expected tRNA genes. The other two tRNA genes, *trnR* and *trnS2*, were identified by hand folding unassigned sequence at the appropriate sites and verified by alignment of the conserved stems and anticodon loops. The rRNA gene boundaries were interpreted as the end of a bounding tRNA gene and by alignment with homologous gene sequences from other insect taxa.

Phylogenetic trees were constructed based on alignments of the ten new sequences, together with complete sequences of 14 other dipterans, selected for broad representation across the order. Table 1 lists the taxa used for phylogenetic analysis. Protein coding genes were extracted and translated using the invertebrate mitochondrial genetic code. The inferred amino acid sequences were aligned using ClustalW2 (Larkin et al. 2007). The alignments were transferred to the DNA sequences, and third codon positions removed. The aligned first and second codon positions were then concatenated into NEXUS and MEGA file formats. The large and small ribosomal sequences were also aligned using ClustalW2 and after manual optimization, were concatenated into the NEXUS and MEGA files.

Phylogenetic trees were constructed using MrBayes 3.1 (Ronquist and Huelsenbeck 2003) with the GTR + I + Γ model, run for 1–3 million generations. The model was selected using jModelTest (Posada 2008). Runs were stopped when the standard deviation of split frequencies fell below 0.005. Neighbor joining trees were constructed using MEGA4 (Tamura et al. 2007).

Table 1
List of Dipteran Taxa Included in This Study

Suborder	Infraorder	Family	Species	Accession	Reference
Nematocera	Tipulamorpha	Tipulidae	*Tipula abdominalis*	JN861743	This study
	Ptychopteromorpha	Ptychopteridae	*Ptychoptera* sp.	JN861744	This study
			Bittacomorphella fenderiana	JN861745	This study
		Tanyderidae	*Protoplasma fitchii*	JN861746	This study
	Bibionomorpha	Pachyneuridae	*Cramptonomyia spenceri*	JN861747	This study
		Keroplatidae	*Arachnocampa flava*	JN861748	This study
		Sciaridae	*Bradysia amoena*	GQ387652	Beckenbach and Joy 2009
		Cecidomyiidae	*Mayetiola destructor*	GQ387648	Beckenbach and Joy 2009
			Rhopalomyia pomum	GQ387649	Beckenbach and Joy 2009
	Culicomorpha	Chironomidae	*Chironomus tepperi*	JN861749	This study
		Ceratopogonidae	*Culicoides arakawai*	NC_009809	Matsumoto Y, Yanase T, Tshuda T, Noda H, unpublished data
		Culicidae	*Anopheles gambiae*	NC_002084	Beard et al. 1993
			Aedes albopictus	NC_006817	Ho C-M, Chang H-P, Liu Y-M, unpublished data
	Psychodomorpha	Trichoceridae	*Trichocera bimaculata*	JN861750	This study
			Paracladura trichoptera	JN861751	This study
		Anisopodidae	*Sylvicola fenestralis*	JN861752	This study
Brachycera	Tabanomorpha	Tabanidae	*Cydistomyia duplonotata*	NC_008756	Cameron, Lambkin, et al. 2007
	Asilomorpha	Nemestrinidae	*Trichophthalma punctata*	NC_008755	Cameron, Lambkin, et al. 2007
	Muscomorpha	Syrphidae	*Simosyrphus grandicornis*	NC_008754	Cameron, Lambkin, et al. 2007
		Muscidae	*Haematobia irritans*	NC_007102	Lessinger AC, Oliveira MT, Barau JG, Feijao PC, Neiva LS, da Rosa AC, Abreu CF, unpublished data
		Calliphoridae	*Cochliomyia hominivorax*	NC_002660	Lessinger et al. 2000
		Oestridae	*Dermatobia hominis*	NC_006378	Azeredo-Espin AML, Junqueira ACM, Lessinger AC, Lyra ML, Torres TT, unpublished data
		Tephritidae	*Ceratitis capitata*	NC_000857	Spanos et al. 2000
		Drosophilidae	*Drosophila melanogaster*	NC_001709	Lewis et al. 1995
Order Mecoptera		Nannochoristidae	*Microchorista philpotti*	HQ696580	Beckenbach 2011
		Boreidae	*Boreus elegans*	NC_015119	Beckenbach 2011
		Bittacidae	*Bittacus pilicornis*	NC_015118	Beckenbach 2011

Note.—Infraorder assignments are based on Wood and Borkent (1989).

Results and Discussion

General Features of the Genomes

The mitochondrial genomes of the Nematocera sequenced in this study are circular, and mostly typical of other insect genomes. Some general characteristics of the genomes are given in Table 2. Annotation of these sequences is given in supplementary tables S1–S10, Supplementary Material Online. The genomes range in size from 15,214 bp in *Ptychoptera* to about 18,600 bp in *Bittacomorphella*, both in the Ptychopteridae. Most of the size variation is due to differences in the control region, although some of the genomes have additional noncoding regions within the coding region. The control region in *Ptychoptera* is about 369 bp (depending on the exact start of the *rrnS* gene); in *Bittacomorphella*, it is about 3.7 kb.

All of the genomes examined here show base composition biases as is usually observed in insect mitochondrial genomes. The A + T content of dipteran coding region ranges from about 73% in *Trichophthalma* and *Trichocera*, to about 83% in the cecidomyiids, *Mayetiola* and *Rhopalomyia*, with a mean of 76.7% (Table 2). A + T content of the N-strand genes, which includes four of the seven NADH dehydrogenase complex genes, is about 3% higher than for the J-strand genes. This result is consistent across all sequences

Table 2
Characteristics of Dipteran and Mecopteran Mitochondrial Genomes

			A + T Content (%)			Control Region		
	Size (bp)	Genome Arrangement[a]	J-Strand	N-Strand	Coding	Size (bp)	Repeats?	%A + T
Tipula	>14,566	A	72.1	75.7	74.3	na	?	na
Ptychoptera	15,214	A	73.2	76.4	75.1	369	no	94.0
Bittacomorphella	~18,600	A	74.0	77.2	75.9	~3,700	3+ (180 bp)	87.7
Protoplasma	16,154	A	73.7	75.7	75.4	1,255	4+ (197 bp)	92.0
Cramptonomyia	16,274	A	71.4	74.8	74.0	1,069	3+ (181 bp)	90.6
Arachnocampa	16,923	*trnE* inv	77.8	80.6	79.7	1,841	4+ (219 bp)	93.3
Bradysia	>14,000	tRNAs inv, trans	74.7	78.0	77.2	na	?	na
Mayetiola	14,759	tRNAs inv, trans	81.6	83.1	82.9	604	no	90.9
Rhaopalomyia	14,503	tRNAs inv, trans	82.9	84.4	84.0	363	no	94.2
Chironomus	15,652	A	72.9	76.5	75.4	535	no	93.3
Culicoides	18,135	A	72.4	75.6	75.1	1,421	5+ (170 bp)	90.6
Anopheles	15,363	tRNAs inv, trans	74.7	77.9	76.6	521	no	94.2
Aedes	16,655	tRNAs inv, trans	75.9	78.4	77.6	1,775	3+ (190 bp)	91.6
Trichocera	16,140	A	70.8	74.5	73.4	1,048	no	89.1
Paracladura	16,143	Extensive trans	74.8	78.2	76.8	904	6 (10–11 bp)	86.9
Sylvicola	16,234	A	73.0	76.2	75.1	1,232	5 (131 bp)	86.0
Cydistomyia	16,247	A	74.1	77.8	76.2	1,378	no	92.6
Trichophthalma	16,396	A	70.3	74.4	72.9	1,599	2+ (227 bp)	81.6
Simosyrphus	16,141	A	77.1	81.4	79.5	1,129	no	91.8
Haematobia	16,078	A	76.0	80.2	78.1	1,261	no	89.5
Cochliomyia	16,022	A	73.1	77.3	75.4	1,177	no	90.7
Dermatobia	16,360	A	74.0	77.2	76.2	1,547	no	91.4
Ceratitis	15,980	A	73.9	78.2	76.2	1,006	no	91.2
Drosophila	19,517	A	75.8	79.3	77.8	4,603	2+ (340), 4+(464)	95.6
Microchorista	>19,092	A	71.1	74.5	73.3	na	?	na
Boreus	16,803	A	77.5	80.6	79.2	1,970	3+ (239 bp)	91.8
Bittacus	15,842	A	70.3	74.0	72.3	1,059	no	83.6

[a] A = ancestral arrangement; inv = inversion; trans = translocation; na = not available; no = not present; ? = unknown.

and probably reflects differences in amino acid content, as well as the well-known strand biases.

Most of the nematoceran sequences retain the ancestral Dipteran gene arrangement. This observation is notable as rearrangements of tRNA genes have been found in mosquitoes (Beard et al. 1993; Mitchell et al. 1993), gall midges, and sciarid midges (Beckenbach and Joy, 2009). Only two of the sequences in this study have rearrangements. *Arachnocampa* (Keroplatidae) has an inversion of the *trnE* gene. *Paracladura* (Trichoceridae) has extensive rearrangements involving major genes as well as tRNA genes and is examined in detail below. The other representative of this family, *Trichocera*, retains the ancestral dipteran gene arrangement.

In the *Chironomus* sequence, *trnW* and *trnC* do not overlap. These genes, coded on opposite strands, overlap in the ancestral gene arrangement by seven residues, comprising the 3′ ends of both amino acyl stems. While this change is not a gene rearrangement, the condition in this sequence required a duplication of at least seven residues.

Transcription Termination Factor Binding Sites

Five primary transcripts have been identified and mapped in *Drosophila melanogaster* (Berthier et al. 1986). The approximate positions and extent of these transcripts are depicted in Figure 1. In the typical insect mitochondrial genome, there are two sites where blocks of genes coded on different strands meet at their downstream ends. These sites are indicated in Figure 1 by vertical arrows. Alignments of the sequences of these two regions are shown in Figure 2 for representative Diptera and Mecoptera. In *D. melanogaster*, 16 bp noncoding sequences having significant sequence similarity are present at both sites (Fig. 2). These sequences have been shown to be binding sites for a bidirectional transcription termination factor, DmTTF (Roberti et al. 2003). Binding of DmTTF has been shown to attenuate transcription in both directions in this species, reducing the production of antisense RNA in each direction beyond those sites (Roberti et al. 2006).

Examination of the first site, between *trnE* and *trnF*, where primary transcripts labeled A and D in Figure 1 meet, show that this binding site is not completely conserved across Diptera and is absent from the Mecoptera (Fig. 2*A*). It is absent as well from other insect orders (Beckenbach and Stewart 2009). Sequences similar to the DmTTF binding site are present in all of the Brachycera and some of the Nematocera but is notably absent from

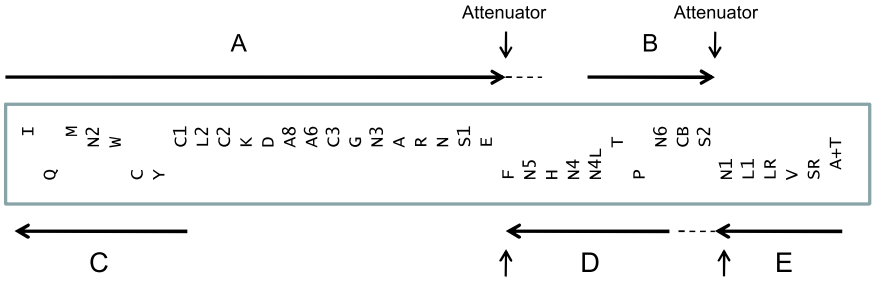

Fig. 1.—Transcription of the mitochondrial genome of *Drosophila melanogaster* (after Berthier et al. 1986). Horizontal arrows indicate the extent of the primary transcripts. Vertical arrows indicate the positions of bidirectional attenuator sequences (Roberti et al. 2003). The short-dashed extensions indicate possible "bleed through" beyond the attenuator sequences.

the mosquitoes. All of the mosquito sequences determined to date have an inversion of the *trnS1* gene, placing it on the N-strand, and requiring it to be transcribed as part of transcript D. The *trnE* gene is not inverted in these sequences but retains its usual position on the J-strand, between the two N-strand genes *trnS1* and *trnF*. It seems likely that the loss of the transcription termination–binding site was a necessary prerequisite for the tRNA gene inversion in mosquitoes.

This binding site is absent from one of the winter crane fly species, *Paracladura*, but present in the other, *Trichocera* (Fig. 2*A*). The *Arachnocampa* sequence is a special case and is omitted from Figure 2. In this species, the *trnE* gene is inverted. Thus transcript D must extend beyond *trnF* to include this gene. A 35 bp noncoding region separates the J-strand gene *trnS1* from the N-strand gene *trnE* in this species, but there is little sequence similarity with the DmTTF binding site sequence. It is evident that this binding site has a function in many Diptera, but is dispensable.

The second DmTTF binding site, between *trnS2* and *nad1*, is more widely conserved. Similar noncoding sequences are present at this site in other insect orders (Cameron and Whiting 2008; Beckenbach and Stewart 2009). All of the sequences determined in this study have a sequence of about the same length and with significant similarity to the DmTTF binding site (Fig. 2*B*). This site has been implicated in the regulation of transcription of the rRNA cassette, transcript E (Fig. 1).

The sequence of *Paracladura* has undergone extensive rearrangement of major and minor genes, as will be detailed below. Among the rearrangements are two that are relevant to this part of the discussion. First, the *trnS2* gene is no longer present between the *cytb* and *nad1* genes. The sequence shown in Figure 2*B* includes part of the *cytb* gene. Although there appears to be some sequence similarity to the DmTTF binding site, its function as a binding site seems doubtful. The other major rearrangement of interest here is that the two rRNA genes have been transposed from

	A	B
	Glu-> <-Phe	Ser-> <-Nad1
Tipula	TATAAATTACTATAATTTATTACGTAAATATATTT-ATTCAAA	ATTAACT---TTACTAATATTTATGAT-------TTAAAATAA
Ptychoptera	TATAAATTACTAATTTTAATTATTTAACT-------GTTTAAA	ATTAGCT---TTACTAAAATTAATTCA-------CTACAATAA
Bittacomorphella	TATAAATTACTATAAATAATTATTTAATT-------ATTTAAA	TTAACTT---ATACTAAATTTTATTAT-------TTAAATAAA
Protoplasma	TATAAATTACTTAAATTAATTATTTAAT--------ATTCAAA	TTAACTT---ATACTAAATTTTATTAT-------TTAAATAAA
Cramptonomyia	TATAAATTACTAAAATAAATTAACTAAT--------ATTTAAA	TTAACTT---ATACTAATATTAATTCT-------CTAAATTAA
Chironomus	TATAAAATTCAACAATACTTATTAAATTTTAATGATATTTAAG	TTAACTT--TATACTAATTTTAATTA--------TTAAAAAAT
Culicoides	TATAAATTGTTACTATAATTAATTCTTTT-------ATTTAAA	ATTGATT...TTACTATAAATTATTCA-------TAAAATTAA
Anopheles	TATAAAT----------ATTTAAA	TTAATTT---ATACTAAATTTTATTCA-------TTAAAATAA
Aedes	TATAAAT----------ATTTAAA	TTATTTA---ATACTAAAAATTATTCA-------TTAAAATAA
Trichocera	TATAAGTTACTATATTATATTACTTAAAT-------ATTTAAA	TTAACTT---TTACTACTTAATATTACCTAAAA-TTAAATTAA
Paracladura	TAATAATTTT----------ATTTAAG	AGATACT---TTACTAAAATAATAACATATTTAATTAAATATT
Sylvicola	TATAAGTTACTAAATTTTATTATTAAATTT-...--ATTCAAA	ATTAACT---TTACTAAAATATATTCATTAAT--TTAAACTAT
Cydistomyia	TATAAATTACTAAATTAAATTCATT----------ACTTAAA	ATTAGCT---TTACTAATATTTATTCATAA----CATTAATGA
Trichophthalma	TATAAGTTACTCTTATTAATTATCTAAT--------ATTCAAA	ATTAACT---TTACTAAATTTAATTCA-------CAAGATTAA
Simosyrphus	TATAAATATTTACTAATTTTAGTTATTTAATT----ATTTAAA	ATTAACT---TTACTATAATTAATTCA-------TTAAACAAA
Haemotobia	TATAAATTACTAAAAAATATTCACTAT---------ATTCAAA	ATTAACTTATTTACTAAAAAATATTCA-------CTATAATAA
Cochliomyia	TATAAATTACTAAAATTAATTCATTAT---------ATTCAAA	ATTAACT---TTACTAAAATAAATTCA-------TTAAATTAA
Dermatobia	TATAAATTACTAAAAAAAATTCATTAT---------ATTCAAA	ATTAACT---TTACTAAAATCAATTCA-------CTATATTAA
Ceratitis	TATAAATTACTAAAAATAATTAACTAT---------ATCTAAG	ATTGACT---TTACTAAAATTAATTAA-------CCACATATA
Drosophila	TATAAATTACTAAAATTAATTCACTAT---------ATCCAAA	ATTAACTT--TTACTAAAAAAAATTCA-------CTATAATAA
Microchorista	TATAAAT----------ATTTAAA	ATTTACT---TTACAAAAAAAGATTCA-------CTATAAATT
Boreus	TATAAAT----------ATTCAAA	ATTAATT---TTACTAAAATTAATTCA-------CTATAATAA
Bittacus	TATAGAT----------ATTCAAA	ATTAACT---TTACTAAATAAAATTCA--------TATAATAA

Fig. 2.—Sequence alignments of the two sites where primary transcripts from opposite strands meet. Due to a gene rearrangement, the junction in *Paracladura* (part *B*) is *cytb–nad1*, rather than *trnS2–nad1*. In *Sylvicola* (part *A*), some additional noncoding residues have been removed.

```
                                    aaaGUUUUAUUUUGGCUUAAAAAUUUGUUAUUAAUUUGAUUUA--UAUGUAAA
Drosophila          ATCAAAAATAAAAATTTAAAGTTTTATTTTGGCTTAAAAATTTGTTATTAATTTGATTTA--TATGTAAA
Ptychoptera         ..A.TTT.ATTT.TA.A..G........C.T.......GT...A..TA..T...AT....--C.......
Bittacomorphella    TATTTGT.A.T.T.A.A..G........C.T............A..TA..G...AT....--C.......
Protoplasma         T.TT.T.TAG.T..AAA...........C.T.........A..TA.TA......T....TA-........
Cramptonomyia       .ATTTT.TAT.TTT.AA.............T...........A..T...T...A.....--C...A...
Arachnocampa        ..TGT...ATTTT.A.A.............TA...........A.A....T.AAT....TA-C.......
Chironomus          T.T.TTT.A.....GGG..G....GT....TTTAA.TGTT..GT..TAA.T..ACTAA..T-..AT..TT
Trichocera          T.ATT..T...TT...A..T..........T..A.........A..T.........G....--C.......
Paracladura         T.ATTTTT.T....AAA.TT..........AA.AA..G.T...TA.T..T..A.T....TA-.......G
Sylvicola           T.A.T.GT.TT..T.A.....A......C.TT.....T..C..A..GAA....AAT....--C.......
Cydistomyia         TAATT..T....TT...............T.......G....A..G.......AT....--C.......
Trichophthalma      CCT.TCT.ATT.CTA.A...........C.T.....T.TT....C.TA..C.A.AT....--C....T..
Simosyrphus         TAA..T.T...TTT.A.............A.........A..TA.TA.T.A..T..A.TA-........
Haematobia          ..T..T...T...T.A.....................TT.....A......G...A.....--........
Cochliomyia         T.T.T.GT.G.TGG.A...................GT......A..........A.....--C.......
Dermatobia          TATTTT.T....TTAA.....................T......A..........A.....--........
Ceratitis           T.TT.TTT.T.TTTA.....................T..T......G.....A.AT....--........
Boreus              ..T..TTT..TTT.A.A...A.......C.T.T...G......A..........TTA...--........
Bittacus            T.T.TT.T.TTTT.A.A.T...............ATT..G...A..T...G..CAT....--C......G
Microchorista       TGGG.TTT.GG.G.A.A..........CC.T......G.TA..TA.T..T.G..T....TCAC.....C.
```

Fig. 3.—N-strand sequence of the junction between the A + T rich region and the 5′ end of *rrnS* genes in Diptera and Mecoptera. The top line shows the 5′ end of the *Drosophila melanogaster* 12S rRNA.

their usual position upstream from the *nad1* gene. There is no evidence of sequence similar to the DmTTF binding site downstream of the *rrnS–rrnL* cassette in its new position, and there are few, if any, noncoding residues in this region.

5′ End of the Small Ribosomal Subunit

Annotation of the 5′ end of the *rrnS* gene in insect mitochondrial sequences has always been somewhat arbitrary (Clary and Wolstenholme 1985). The junction between the A + T rich region and the *rrnS* gene of representative Diptera and Mecoptera are shown in Figure 3. The 5′ end of *rrnS* of *D. melanogaster* has been mapped by circularization and reverse transcriptase PCR (Stewart and Beckenbach 2009). The start of the rRNA sequence is indicated in the top line of the alignment. The technique does not allow us to distinguish whether any of the first three residues, shown as lower case (aaa), are part of the gene or derived from the poly-A tail and attached to the 5′ end during the circularization process. The alignment in Figure 3 represents more than 250 Myr of evolution, and the relatively high degree of conservation across Diptera and Mecoptera suggests that the start of *rrnS* is AARGUUUU, as observed in *Drosophila*.

Noncoding Regions

Most of the genomes determined in this study are extremely compact, with few noncoding sequences outside of the control region. Several of the sequences have insertions ranging from 99 to 210 bp, for which no coding role is apparent. The *Arachnocampa* sequence includes a 140 bp insert between the *trnI* and *trnQ* genes. *Cramptonomyia* has a 113 bp insert between *nad6* and *cytb*, as well as several smaller inserts elsewhere in the coding region. *Sylvicola* has a 99 bp insert between *trnE* and *trnF*. *Trichocera* has a 185 bp insert between *trnR* and *trnN*. Finally, *Paracladura* has a 210 bp insert between *nad6* and *trnS2*. In this sequence, the *cytb* gene, which is normally located between these two genes, has been moved to another location. It is possible that this insert represents the remnant of a *pseudo-cytb*, but if so, it is no longer recognizable.

The A + T Rich Regions of Nematocera

Four of the eight sequences, where complete A + T rich regions were determined, were relatively small, ranging from 369 bp in *Ptychoptera* to 1,048 bp in *Trichocera* (Table 2). There is no evidence of repeat motifs in three of these sequences. *Paracladura* has a short 10–11 bp sequence (CCTTTTTTGG or CCATTTTTTGG) tandemly repeated six times. Five of the sequences include larger tandem repeats present in three or more copies. *Sylvicola* has a 131 bp sequence repeated five times. *Cramptonomyia* has a 181 bp sequence present in three perfect copies, with a partial fourth. In *Protoplasma*, there is a tandem repeat of a 197 bp sequence, present in four copies with a partial fifth. *Arachnocampa* has four copies of a 219 bp sequence. Finally, *Bittacomorphella*, with the largest control region encountered in this study (about 3.7 kb), has a 180 bp sequence tandemly repeated at least three times. The middle portion of the sequence of the A + T rich region in this species was not determined, in part because of its size and the presence of repeat sequences.

Rearrangement in a Winter Crane Fly Genome

A majority of Diptera mitochondrial sequences share the gene arrangement first encountered in *D. yakuba* and subsequently observed in many other insect orders. The few exceptions are tRNA transpositions or inversions found in mosquitoes (Beard et al. 1993; Mitchell et al. 1993), and

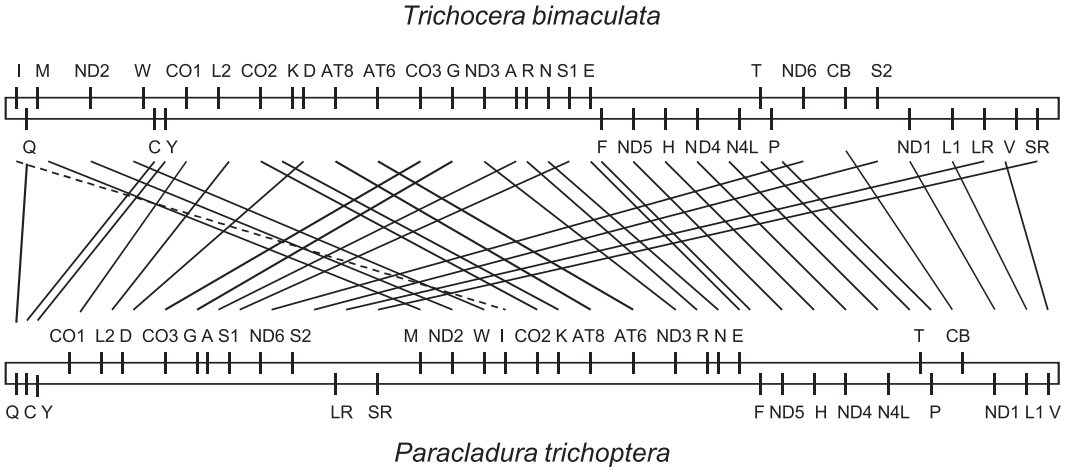

Fig. 4.—Gene arrangements in two species of winter crane fly (Family Trichoceridae). *Trichocera bimaculata* retains the ancestral dipteran arrangement. *Paracladura trichoptera* has undergone extensive rearrangement. Genes shown above each rectangle are transcribed from the majority strand. Those below the rectangles are transcribed from the minority strand.

in gall and sciarid midges (Beckenbach and Joy 2009). The finding of extensive rearrangement including both tRNA and major gene sequences in a winter cranefly, *P. trichoptera*, is unusual, particularly since another winter crane fly, *T. bimaculata* retains the widespread ancestral dipteran arrangement.

A comparison of the arrangements present in these two trichocerids is shown in Figure 4. The rearrangements in *Paracladura* appear to fall into two main groups. Within each group, both the ancestral gene order and coding direction are maintained. The only exception is a transposition of the *trnI* gene from its usual position adjacent to the control region, to a position between the *trnW* and *cox2* genes. The overall pattern depicted in Figure 4 suggests a simple model to explain all of the rearrangement, except for the *trnI* transposition. The model is shown in Figure 5. The approximate positions of the primary transcripts (from Fig. 1) are included in this figure. For simplicity, the tRNA genes are omitted, except for the N-stand tRNAs derived from transcript C. For this model, we assume that a tandem duplication of the entire genome occurred, as depicted in Figure 5*B*. It is also necessary to assume that all genes in both copies of the duplicated genome were fully functional. Evidence has been presented that genes in a large duplication of coding region in a scorpion fly (Order Mecoptera) were initially functional (Beckenbach 2011). We assume that one copy of each gene loses function and is eventually lost through deletions. This model, complete genome duplication followed by loss of one copy of each gene, can account for nearly all of the gene rearrangement in *Paracladura*. If this model is correct, we can make some inferences about the process of elimination of duplicate gene copies.

The most commonly invoked model for gene rearrangement is the duplication/random loss model (Boore 2000). If the loss of one copy of each gene is random, we would expect about half of the genes from copy 1 to be retained and the other half retained from copy 2. With 14 of the genes retained from copy 1 and the other 23 genes retained from copy 2 (Figs. 4 and 5), random loss cannot be rejected ($\chi^2 = 2.19$, 1 degrees of freedom, not significant).

Random loss of genes requires gene-by-gene loss of function. A case can be made for nonrandom loss of some of the genes. In order to function, the region containing the gene must be transcribed. Because there are evidently multiple primary transcripts in the *Drosophila* mitochondrial genome, loss of an initiator would inactivate an entire block of genes (Figs. 1 and 5). Transcript A, for example, includes all J-strand genes from *trnI* to *trnE* in the *Drosophila* mitochondrial genome, a total of 19 genes. In *Paracladura*, seven of these genes are present in the first block from copy 1 and 12 are in the second block from copy 2. Both regions must be transcribed and initiators for both transcripts A and A′ (Fig. 5*B*) must be retained. Random gene-by-gene loss of function and removal appears likely.

In contrast, transcript D includes six N-strand genes, from *trnP* to *trnF*. In *Paracladura*, all six genes are derived from copy 2. If gene loss is random, the probability that all six genes are lost from the same copy is $2\ (1/2)^6 = 0.031$. Berthier et al. (1986) hypothesized that the initiator for the transcript responsible for function of these six genes in the *Drosophila* mitochondrial genome is in either the *nad6* or *cytb* gene. The detection of antisense RNA corresponding to the *nad6* gene in their study (transcripts q and r in their Fig. 3) suggests that the initiator is actually in *cytb*. Loss of the transcription initiator for transcript D from copy 1 in *Paracladura* would inactivate all six genes simultaneously. The *cytb* gene, but not the *nad6* gene, is upstream from the N-strand *trnP* to *trnF* block in *Paracladura* (transcript D′, Fig. 5*B*).

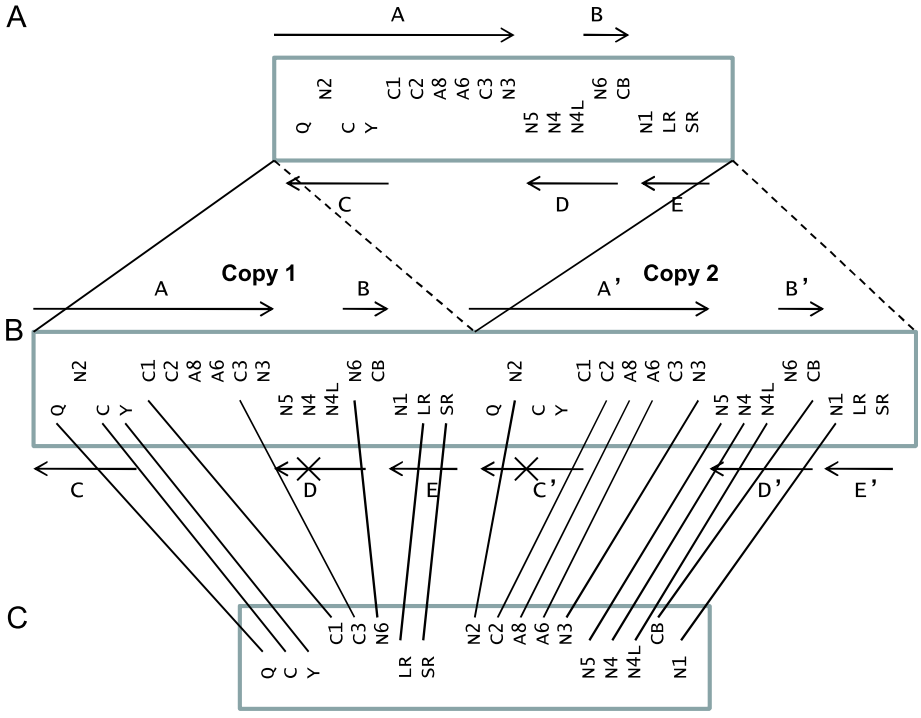

Fig. 5.—Hypothesis to explain the rearrangements observed in *Paracladura*. (*A*) Ancestral arrangement; (*B*) Hypothetical intermediate after complete genome duplication; (*C*) Gene arrangement in *Paracladura*. Most of the tRNA genes are omitted for simplicity. Horizontal arrows in parts *A* and *B* show the probable positions of primary transcripts. Transcripts D and C′ (part *B*) have no apparent coding function in *Paracladura* as indicated by crosses on each arrow.

A second example may be provided by the N-strand tRNA genes *trnQ*, *trnC*, and *trnY*, derived from primary transcript C (Figs. 1 and 5). Berthier et al. (1986) hypothesized an initiator in the *cox1* gene. If their interpretation is correct, the removal of the *cox1* gene from copy 2 (Fig. 5*B*) removes the initiator for primary transcript C′. Since there are only three genes involved (or four, including *cox1*) there is insufficient power for a statistical test. Thus the position of these genes is consistent with either model, random gene-by-gene inactivation or loss of the transcription initiator.

Lavrov et al. (2002) argued that rearrangements they observed in the mitochondrial sequences of two species of millipedes occurred through a similar mechanism: complete genome duplication followed by loss of transcription promotors. Their model provides a very simple mechanism for bringing together genes with a common transcriptional polarity. They assumed the presence of only two promotors, one for each strand, as has been demonstrated in vertebrates (Taanman 1999). If the basic mechanism of transcription in basal arthropods follows the *Drosophila* model (Fig. 1 and 5*A*), the rearrangements in millipedes would appear to require the loss of seven promotors, retaining only promotors for transcripts A, E′, and C (Fig. 5*B*). The promotor for transcript C is required for the *trnC* gene and provides a reasonable explanation for its exceptional position as the only N-strand gene present in the J-strand coding block.

A Mitochondrial Phylogeny of Diptera

Traditionally, the order Diptera has been divided into two suborders, Nematocera ("thread horn") and Brachycera ("short horn"), based partly on the structure of the antennae. While the Brachycera is generally believed to be monophyletic, the Nematocera is almost certainly paraphyletic to the Brachycera. That is, the Brachycera arose from within the Nematocera and has as its sister only part of the Nematocera. To avoid this problem, there is a recent proposal to raise the infraorders of the Nematocera to suborder status (Amorim and Yeates 2006). Although this proposal eliminates the need for formal recognition of Nematocera, it may create other problems. In particular, the number and composition of nematoceran infraorders has long been subject to debate, and there remains the possibility that one of the infraorders is itself paraphyletic to the Brachycera. Resolution of these issues requires a robust phylogeny that includes representatives from most of the nematoceran infraorders.

Cameron, Lambkin, et al. (2007) developed a phylogeny of some Brachycera, based on complete mitochondrial genome sequences. The major advantage of using complete sequences is that it makes available large amounts of data. Their analysis proved consistent with well-established relationships within the Brachycera. The Brachycera originated in the Jurassic and underwent two radiations (Wiegmann et al.

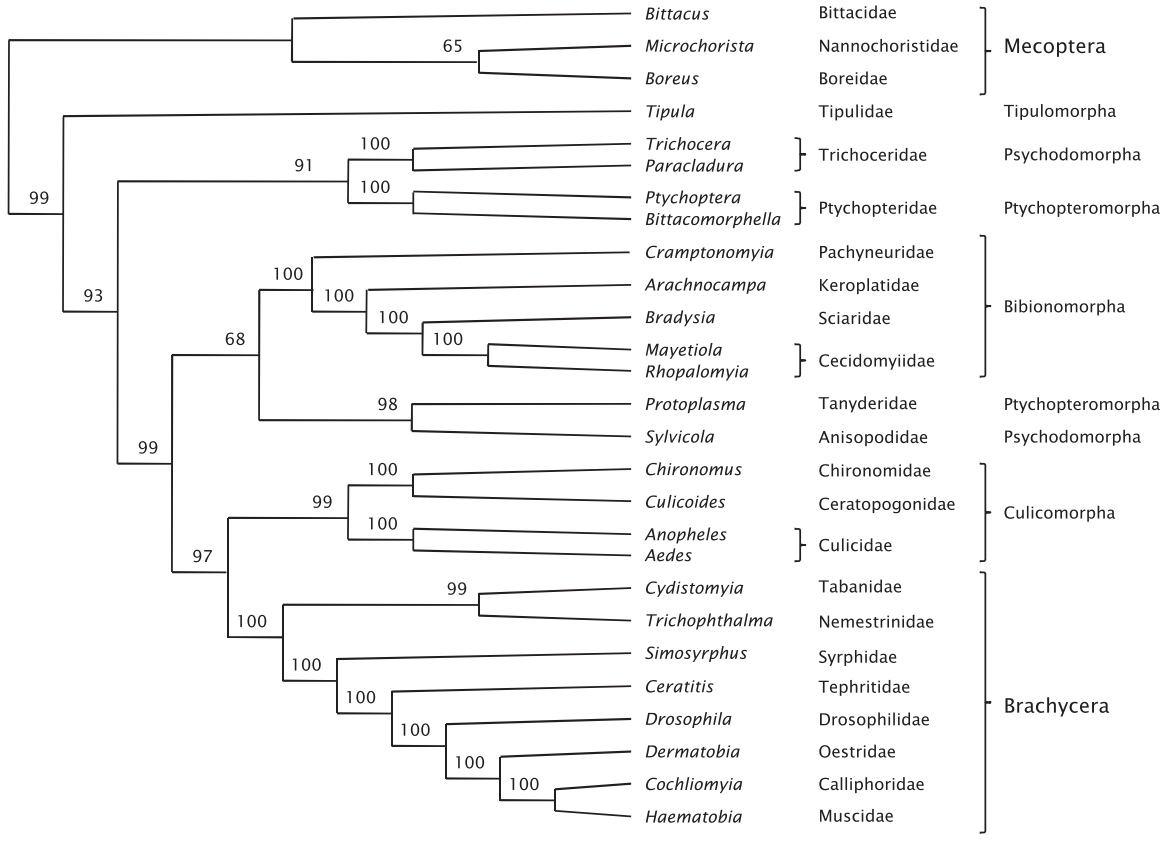

Fig. 6.—A mitochondrial phylogenetic tree of major groups of Diptera. The tree is derived from a Bayesian analysis of all major genes, using codon positions 1 and 2 for protein coding genes, and all alignable sites for the ribosomal genes. Numbers above the branches are credibility scores. The tree is rooted with taxa from the related Order Mecoptera (Scorpion flies).

2011). The earlier radiation, between about 180 and 120 Ma, gave rise to the lower ("orthorrhaphous") Brachycera, while a second radiation between 70 and 40 Ma gave rise to the higher flies. At the time of that study (Cameron, Lambkin, et al. 2007), complete mitochondrial sequences were available for only one family of Nematocera, the Culicidae (mosquitoes). The mosquito sequences emerged as a sister to the remainder of the Diptera (i.e., the Brachycera), as expected.

Resolution of the earliest dipteran radiation, which gave rise to most of the nematoceran families between about 280 and 240 Ma, is particularly challenging. We now have complete (or nearly complete) mitochondrial sequences from representatives of 12 nematoceran families, including representatives from five of perhaps seven nematoceran infraorders. A tree based on Bayesian analysis of first and second codon positions of aligned sequences of all protein coding genes, as well as the small and large ribosomal subunits, is given in Figure 6. In Figure 7, a Bayesian tree is shown based on the same data, except that the *nad1-6*, *nad4l*, and *atp8* genes are omitted. These genes are difficult to align, and the likelihood of including many misaligned sites may pose problems for phylogenetic reconstruction (Nardi et al. 2003).

A potential problem for deep molecular phylogenies is the presence of sequences having greatly differing nucleotide content (Jermiin et al. 2004). In the sequences included in this study, the A + T content of the coding regions vary from about 73% to more than 83% (Table 2). The concern is 2-fold. Not only do the very high A + T content sequences represent very long branches, raising the possibility of long-branch attraction, but also the presence of very high A + T content in protein coding genes necessitates an emphasis on A + T rich codons. Long-branch attraction does not require convergence of the sequences (Felsenstein 1978), but the over utilization of only a subset of codons may exacerbate the long branch problem by superimposing convergence on the long branch problem. A neighbor joining tree based on the data set used for the tree in Figure 7 is given in Figure 8, to illustrate the branch length problem. The most extreme base composition bias and long branches are the two gall midge taxa (Cecidomyiidae). These taxa emerge as sisters in all three trees (Figs. 6–8). There is ample evidence from morphology that this result reflects a true sister relationship. There are no other branches long enough to be attracted to the gall midge branch through the artifact of long branch attraction.

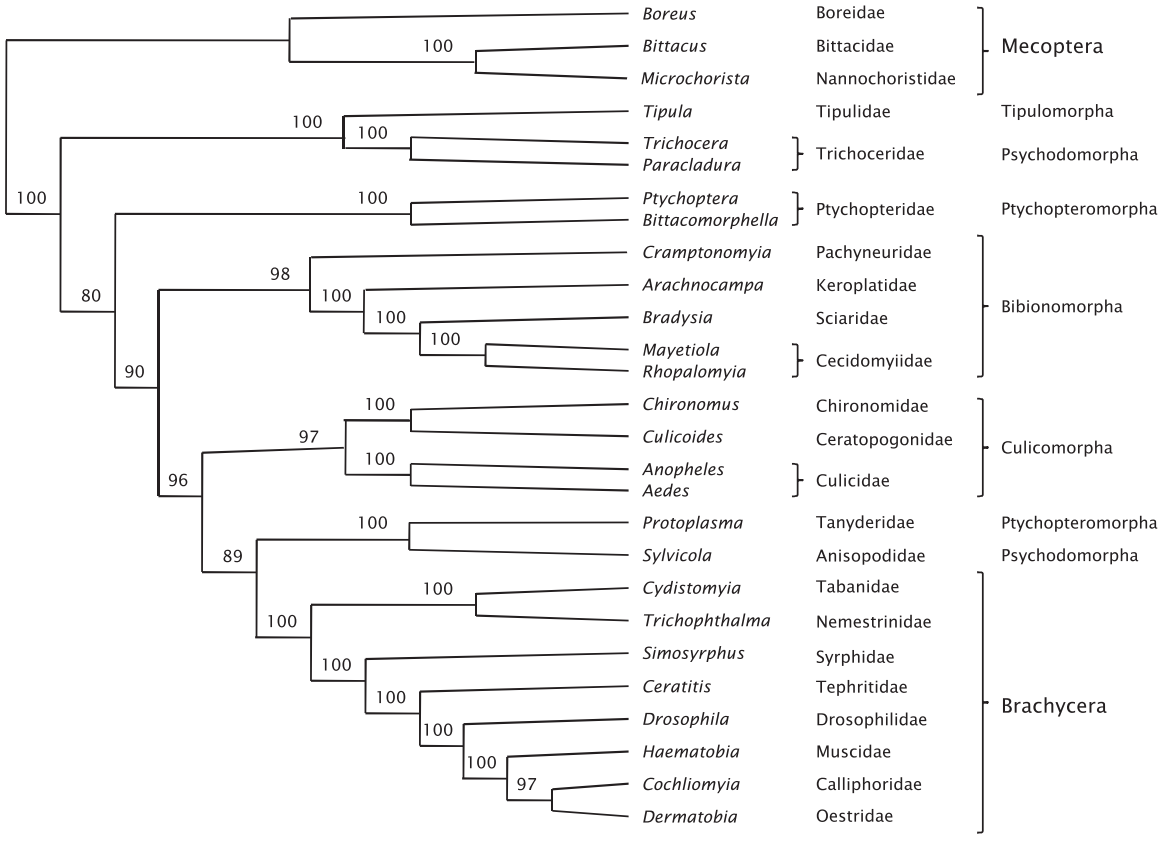

Fig. 7.—A Bayesian mitochondrial tree using codon positions 1 and 2 for *cox1–3*, *cytb*, and *atp6* genes, and all alignable sites for the ribosomal genes. Numbers above the branches are credibility scores. Numbers below the branches are neighbor joining bootstraps. The tree is rooted with the Mecoptera.

These trees give considerable insight into the early diversification of Diptera. The trees are rooted with sequences from representatives of a related order, Mecoptera (scorpion flies). Four of the families are represented in this study by members of two genera: *Ptychoptera* and *Bittacomorphella* in the Ptychopteridae; *Trichocera* and *Paracladura* in the Trichoceridae; *Mayetiola* and *Rhopalomyia* in the Cecidomyiidae; and *Anopheles* and *Aedes* in the Culicidae. In all cases, members of the same family appear as sister taxa, as expected (Figs. 6–8).

Monophyly of Infraorder Culicomorpha, including mosquitoes (Culicidae), biting midges (Ceratopogonidae), and chironomid midges, is well supported. This assemblage has long been recognized as a natural grouping, and the pairing of the Chironomidae and Ceratopogonidae is consistent with their usual placement in the same superfamily or family group (Hennig 1973; Wood and Borkent 1989; Oosterbroek and Courtney 1995).

Monophyly of the Bibionomorpha is also well supported. The families included in this study exhibit the same branching order as is observed based on morphology (Wood and Borkent 1989; Oosterbroek and Courtney 1995). The close relationship between the Sciaridae and Cecidomyiidae is consistent with other genetic evidence. Members of both families undergo elimination of chromosomes from somatic cells during development, use elimination of X chromosomes for sex determination, and display an unusual form of meiosis in males, without chromosome pairing (White 1949). These features have not been found in flies from any other family.

Infraorder Tipulomorpha has been variously defined to include both the Tipulidae, sensu lato (crane flies), and Trichoceridae (winter crane flies) (Hennig 1973; Bertone et al. 2008) or just the Tipulidae, sensu lato (Wood and Borkent 1989). Oosterbroek and Courtney (1995) placed them together in the "higher" Nematocera. Mitochondrial sequence data do not provide a clear resolution of this question. Exclusion of the more variable major genes supports the pairing of these families (Figs. 7 and 8), whereas inclusion of all major genes supports defining an infraorder Tipulomorpha consisting only of the Tipulidae sensu lato (Fig. 6). In either case, the Tipulomorpha emerge as the earliest branch of the Diptera included in this study (Figs. 6 and 7).

Infraorder Ptychopteromorpha was erected to include two families, Ptychopteridae (false and phantom crane flies) and Tanyderidae ("primitive" crane flies) (Wood and Borkent

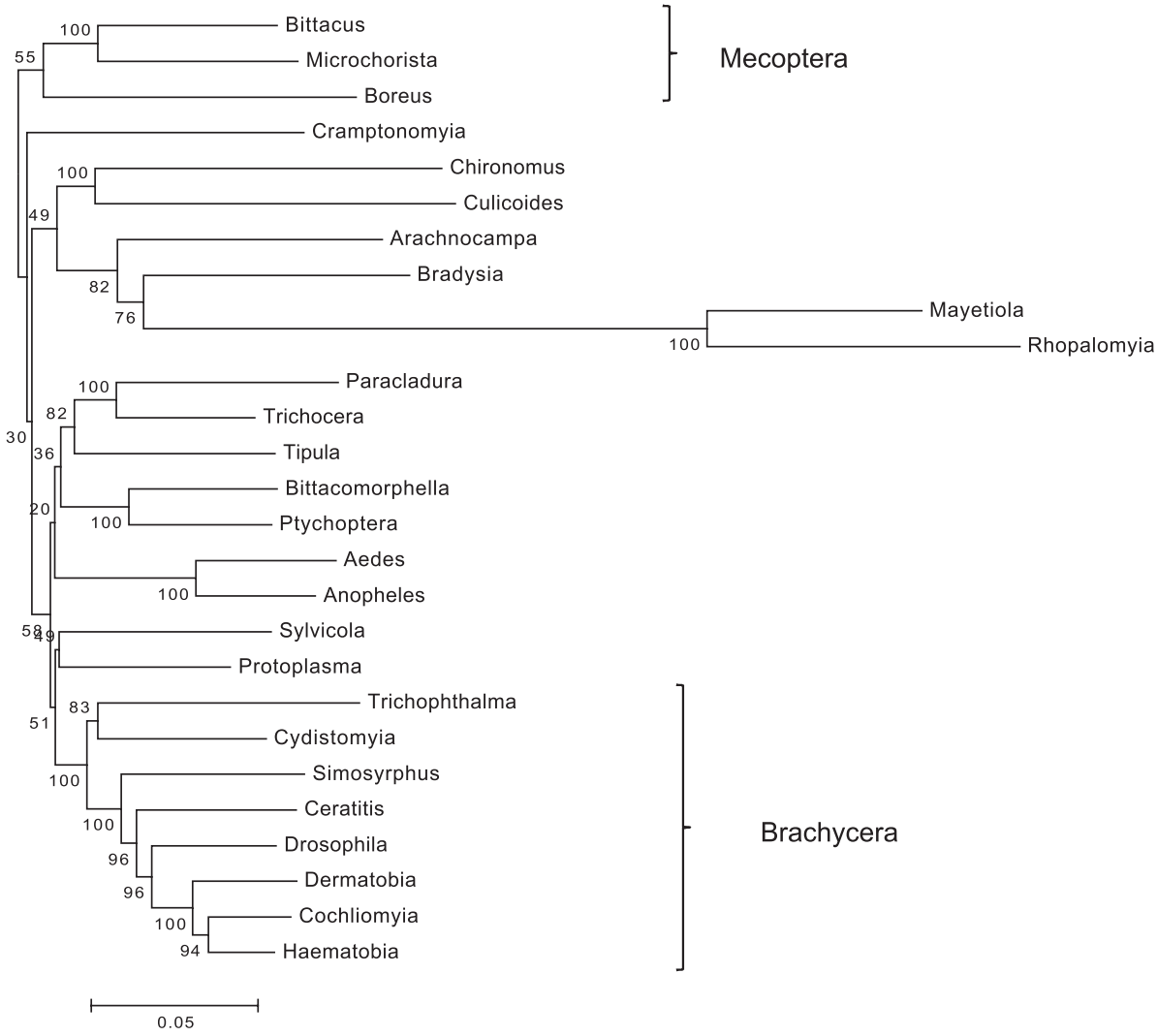

Fig. 8.—Neighbor joining tree using the same data set as Figure 7, showing the branch lengths. Numbers adjacent to each node are bootstraps.

1989). The relationship between the families is supported by a single morphological character, which is absent in some ptychopterids (Oosterbroek and Courtney 1995). Molecular studies have failed to support the placement of the Tanyderidae with the Ptychopteridae (Bertone et al. 2008; Wiegmann et al. 2011). When all genes are included, the mitochondrial sequence data groups the Ptychopteridae with the Trichoceridae, diverging from the rest of the Diptera after the tipulids (Fig. 6). When the more variable mitochondrial genes are excluded, the Ptychopteridae appear on its own branch (Fig. 7).

Some authors include the Anisopodidae (wood gnats) in the Bibionomorpha (Hennig 1973; Bertone et al. 2008; Wiegmann et al. 2011). Wood and Borkent (1989) placed the family in the Psychodomorpha. The placement of Anisopodidae is of particular interest because of morphological similarities of the adults to some Brachycera, suggesting this family as a possible sister to the Brachycera (Woodley 1989; Oosterbroek and Courtney 1995). The mitochondrial trees place the Anisopodidae with the Tanyderidae (Figs. 6–8). The Anisopodidae and Trichoceridae were placed in the infraorder Psychodomorpha by Wood and Borkent (1989). There is no evidence in the mitochondrial trees for this pairing. Unfortunately, there are no complete mitochondrial sequences available for representatives of any other psychodomorph families, and the inclusion of these families with other families of this infraorder has not been widely accepted. The infraorder Psychodamorpha is poorly defined (Bertone et al. 2008).

The origin of the Brachycera has long been subject to debate (Woodley 1989). All trees give strong support for monophyly of this suborder, and confirm that the Nematocera is paraphyletic to the Brachycera. The more restricted data sets give the Anisopodidae + Tanyderidae as sister to

the Brachycera (Figs. 7 and 8), while the inclusion of all gene sequences suggests that the Culicomorpha is the sister (Fig. 6). The former result is more consistent with the findings of other studies.

In general, the use of complete mitochondrial genomes for resolving questions of the early diversification of Diptera shows considerable promise. More complete sampling of the Nematocera and the lower ("orthorrhaphous") Brachycera should help clarify many of the outstanding questions of dipteran phylogeny.

Supplementary Material

Supplementary tables S1–S10 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Funding for this study was provided by a Natural Sciences and Engineering Research Council of Canada Discovery Grant and a National Science Foundation (United States) grant, EF-0334949, Building the Dipteran Tree of Life.

Literature Cited

Amorim DS, Yeates DK. 2006. Pesky gnats: ridding dipteran classification of the Nematocera. Studia Dipterologica 13:3–9.

Beard CB, Hamm DM, Collins FC. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol Biol. 2:103–124.

Beckenbach AT. 2011. Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. Genome 54:368–376.

Beckenbach AT, Joy JB. 2009. Evolution of the mitochondrial genomes of gall midges (Diptera: Cecidomyiidae): rearrangement and severe truncation of tRNA genes. Genome Biol Evol. 2009:278–287.

Beckenbach AT, Stewart JB. 2009. Insect mitochondrial genomics 3: the complete mitochondrial genome sequences of representatives from two neuropteroid orders: a dobsonfly (order Megaloptera) and a giant lacewing and an owlfly (order Neuroptera). Genome 52:31–38.

Berthier F, Renaud M, Alziari S, Durand R. 1986. RNA mapping on *Drosophila* mitochondrial DNA: precursors and template strands. Nucleic Acids Res. 14:4519–4533.

Bertone MA, Courtney GW, Wiegmann BM. 2008. Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes. Syst Entomol. 33:668–687.

Boore JL. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27:1767–1780.

Boore JL. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of Deuterostome animals. In: Sankoff D, Nadeau JH, editors. Comparative genomics. Dordrecht (The Netherlands): Kluwer Academic Publishers. p. 133–147.

Boore JL, Lavrov DV, Brown WM. 1998. Gene translocation links insects and crustaceans. Nature 392:667–668.

Cameron SL, Beckenbach AT, Dowton M, Whiting MF. 2006. Evidence from mitochondrial genomics on interordinal relationships in insects. Arthropod Syst Phylogeny 64:27–34.

Cameron SL, Johnson KP, Whiting MF. 2007. The mitochondrial genome of the screamer louse *Bothriometopus* (Phthiraptera: Ischnocera): effects of extensive gene rearrangements on the evolution of the genome. J Mol Evol. 65:589–604.

Cameron SL, Lambkin CL, Barker SC, Whiting MF. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst Entomol. 32:40–59.

Cameron SL, Whiting MF. 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 408:112–123.

Clary DO, Wolstenholme DR. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J Mol Evol. 22:252–271.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool. 27:401–410.

Hennig W. 1973. Ordnung Diptera (Zweiflügler). In: Helmcke JG, Starck D, Wermuth H, editors. Handbuch der Zoologie, Bd. 4: Arthropoda, 2. Hälfter: Insecta, 2. Aufl., 2 Teil Spezielles. Berlin (Germany): Walter de Gruyter. p. 1–377.

Jermiin LS, Ho SYW, Ababneh F, Robinson J, Larkum AWD. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst Biol. 53:638–643.

Larkin MA, et al. 2007. ClustalW and ClustalX version 2. Bioinformatics 23:2947–2948.

Lavrov DV, Boore JL, Brown WM. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. Mol Biol Evol. 19:163–169.

Lessinger AC, Martins Junqueira AC, Conte FF, Azeredo-Espin AM. 2004. Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. Gene 339:1–6.

Lessinger AC, et al. 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). Insect Mol Biol. 9:521–529.

Lewis DL, Farr CL, Kaguni LS. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol Biol. 4:263–278.

Liu H, Beckenbach AT. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. Mol Phylogenet Evol. 1:41–52.

Lowe TD, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Mitchell SE, Cockburn AF, Seawright JA. 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome 36:1058–1073.

Nardi F, et al. 2003. Hexapod origins: monophyletic or paraphyletic? Science 299:1887–1889.

Oosterbroek P, Courtney G. 1995. Phylogeny of the nematocerous families of Diptera (Insecta). Zool J Linn Soc. 115:267–311.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25:1253–1256.

Roberti M, Bruni F, Polosa PL, Gadaleta MN, Cantatore P. 2006. The *Drosophila* termination factor DmTTF regulates *in vivo* mitochondrial transcription. Nucleic Acids Res. 34:2109–2116.

Roberti M, et al. 2003. DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences in *Drosophila melanogaster* mitochondrial DNA. Nucleic Acids Res. 31:1597–1604.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Shao R, Barker SC. 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol Biol Evol. 20:362–370.

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst. 37: 545–579.

Spanos L, Koutroumbas G, Kotsyfakis M, Lewis C. 2000. The mitochondrial genome of the Mediterranean fruitfly *Ceratitis capitata*. Insect Mol Biol. 9:139–144.

Stewart JB, Beckenbach AT. 2009. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implicatons for the tRNA punctuation model in arthropods. Gene 445:49–57.

Taanman J-W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta 1410:103–123.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4. Mol Biol Evol. 24:1596–1599.

White MJD. 1949. Cytological evidence on the phylogeny and classification of the Diptera. Evolution 3:252–265.

Wiegmann BM, et al. 2011. Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A. 108:5690–5695.

Wood DM, Borkent A. 1989. Phylogeny and classification of the Nematocera. In: McAlpine JF, Wood DM, editors. Manual of Nearctic Diptera Vol. 3. Ottawa (Canada): Agriculture Canada Research Branch. p. 1333–1370.

Woodley NE. 1989. Phylogeny and classification of the "orthorrhaphous" Brachycera. In: McAlpine JF, Wood DM, editors. Manual of Nearctic Diptera Vol. 3. Ottawa (Canada): Agriculture Canada Research Branch. p. 1371–1395.

Associate editor: Gertraud Burger