Insect Molecular Biology (2001) **10**(5), 475–485

Molecular phylogeny of Calyptratae (Diptera: Brachycera): the evolution of 18S and 16S ribosomal rDNAs in higher dipterans and their use in phylogenetic inference

X. Nirmala†, V. Hypša*'‡ and M. Žurovec†'‡

**Institute of Parasitology and †Institute of Entomology, Academy of Sciences of the Czech Republic; and ‡Faculty of Biological Sciences, University of South Bohemia, České Budějovice, Czech Republic*

Abstract

Sequences for nearly complete 18S rRNA and partial 16S rRNA genes were determined for sixteen species representing twelve calyptrate families. Two unique insertions are present in expansion regions of the 18S rDNA in nycteribiids. Alignments containing other dipteran rRNA genes provided good resolution at higher taxonomic level: monophyly of Calyptratae is well supported. While both 16S and 18S rDNA matrices produce unstable topologies within Calyptratae when analysed separately, their combination results in a tree with several robust and well supported nodes. Of three superfamilies recognized in recent classifications, the Hippoboscoidea is well supported by 16S rDNA and by combined matrices. The representatives of Muscoidea, *Musca* sp. and *Antipoda* sp., display a tendency to cluster within Oestroidea. The comparison of secondary structures of two variable regions indicates that Sarcophagidae are related to Calliphoridae rather than to Tachinidae.

Keywords: phylogeny, Calyptratae, ribosomal RNA, parasitic diptera, tsetse flies, RNA secondary structure.

Introduction

Calyptratae are one of the major dipteran groups, and certainly the most diversified within this order (McAlpine, 1989; Yeates & Wiegman, 1999). They include well-known groups such as house-flies, flesh-flies, tsetse-flies, warble-flies, etc. Compared to other dipteran lineages, the calyptrate taxa evolved an impressive variety of life strategies and several unique bionomical features including a bloodsucking habit associated with a viviparous mode of reproduction and numerous cases of larval endoparasitism. Considering the variety of habitats invaded by these parasitic flies, the most fascinating question behind the Calyptratae phylogeny is therefore the origin, evolution and mutual relationships of parasitic lineages. Despite a long history of morphological studies (for review see McAlpine, 1989; Yeates & Wiegman, 1999) and the attention drawn by medically and veterinary important groups (Glossinidae, Oestridae), the phylogeny of higher diptera, and evolution of their diverse life styles remain unclear. The most widely accepted is a McAlpine's classification resting exclusively on morphological studies (McAlpine, 1989; also refer to *Tree of life*: http://phylogeny.arizona.edu/tree/phylogeny.html; Fig. 1). According to this classification, the Calyptratae are divided into three superfamilies, Hippoboscoidea, Muscoidea and Oestroidea.

The hippoboscoid superfamily encompasses several obligatory bloodsucking groups associated with mammals. In addition to the well-known medically important family Glossinidae, the group includes Hippoboscidae, the bloodsucking parasites of birds and mammals, and the two bat-associated families, Nycteribiidae and Streblidae. Their common origin is supported by several morphological and bionomical characters, of which the most conspicuous is a unique mode of larval development, called adenotrophic viviparity. However, the monophyly of this clade was not always accepted (Bequaert, 1954; Wenzel *et al*., 1966; Schlein, 1970; Pollock, 1971), and even the viviparity was considered a convergent feature (Hagan, 1951). In a recent review on evolution and taxonomic distribution of various forms of viviparity and ovoviviparity in Diptera, the authors give many examples of the evolutionary tendencies towards viviparity and determine at least sixty-one independent origins of this reproductive mode within Diptera (Meier *et al*., 1999). This shows that one should be careful not to overestimate the phylogenetic significance of even such apparently derived bionomical features. The remaining calyptrate families are usually treated as two distinct lineages. Those displaying tendencies towards a parasitic life style are classified within the superfamily Oestroidea.

Received 15 December 2000; accepted after revision 25 May 2001. Correspondence: Václav Hypša, Institute of Parasitology, Academy of Sciences of the Czech Republic, Branišovská 31, 370 05 České Budějovice, Czech Republic. Tel.: +420 38 7775441; fax: +420 38 5300388; e-mail: vacatko@paru.cas.cz

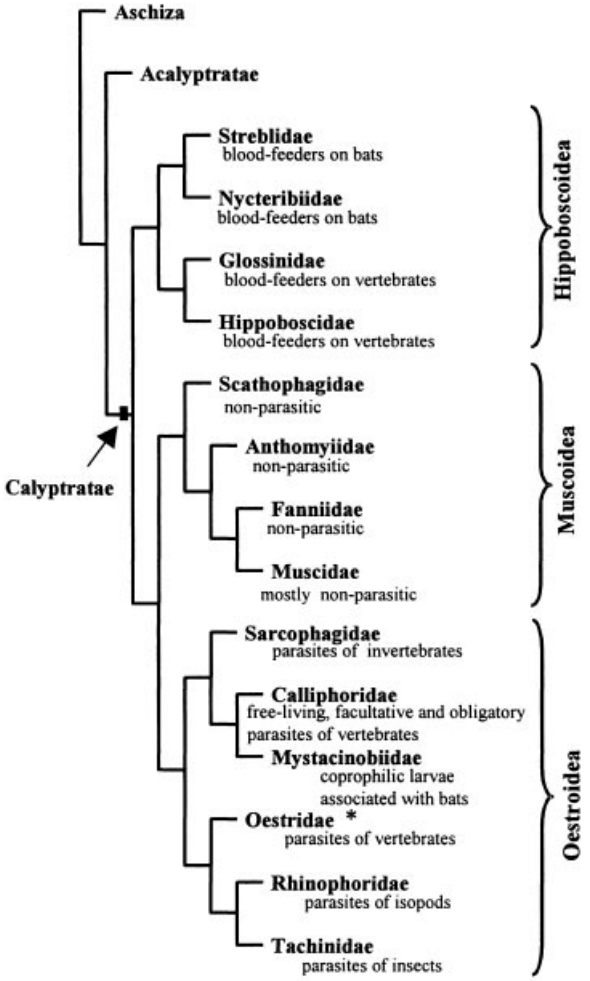

Figure 1. Morphology-based phylogeny of Cyclorrhapha, according to McAlpine (1989).

The host groups and the mode of parasitism of individual oestroid families are shown in Fig. 1. The non-parasitic families Muscidae, Fanniidae, Scathophagidae and Anthomyiidae are classified within the Muscoidea superfamily (Hennig, 1973; McAlpine, 1989; but see Griffiths, 1976). However, while the distinction between hippoboscoids on one hand and all other calyptrates on the other hand is well substantiated, there is no consensus on the monophyly of muscoids and oestroids and the relationships between their families.

In an attempt to add new sources of data, several molecular studies were performed during the last decade without any clear conclusion (Vossbrinck & Friedman, 1989; Carreno & Barta, 1998; Bernasconi *et al.*, 2000). These studies show that molecular investigations of higher Dipteran phylogeny faces two major problems: a limited number of species and families represented by DNA sequences, and the lack of reliable phylogenetic information covering all taxonomic levels within the Calyptratae. Because the radiation of Calyptratae families seems to have occurred within a relatively short period of time (Vossbrinck & Friedman, 1989; Friedrich & Tautz, 1997) it may prove difficult to find a single gene providing an appropriate phylogenetic information at the interfamilial level within Calyptratae. In this study, we analyse the almost complete sequence of the nuclear 18S rRNA gene and a partial sequence of the mitochondrial 16S rRNA gene from sixteen species of Calyptratae, in an attempt to resolve the lineage of the superfamilies. We report here the patterns of their primary and secondary structures, which provide phylogenetic inference within Calyptratae.

Results

Divergence of 18S and 16S rRNA genes in Calyptratae

We obtained nearly complete sequence of 18S rRNA gene for fourteen species of the superfamilies Hippoboscoidea, Oestroidea and Muscoidea (Table 1). The sequence lengths vary from 1839 bp in Calliphoridae to 2053 bp in Nycteribiidae. The sequenced region corresponds to positions 4–1850 of *Drosophila melanogaster* 18S rRNA gene (GenBank accession number M21017; Fig. 2). Alignment of these sequences together with 18S rRNA genes of six other dipterans retrieved from the GenBank (Table 1) reveals that the majority of sequence variation accumulates in the variable regions V2, V4, and V6 while the V1, V3, and V5 expanding regions (Hancock *et al.*, 1988) are highly conserved in terms of length, and produce a stable unequivocal alignment.

In both nycteribiid species, the V4 region contains a 100-bp long insert while the V6 region includes two inserts of 130 and 20 bp, respectively (Fig. 2). Two short inserts (approx. 10 bp) representing complementary strands of extended helix structure are present in the V4 region of *Cuterebra* (Fig. 2). The occurrence of all insertions was confirmed by cloning the PCR product and subsequent sequencing of the clone, and also by direct sequencing of the PCR product.

After the removal of all variable, ambiguously aligned regions, the alignment retains 1723 positions. Of the 340 variable characters, 146 are parsimony-informative. The distances between the sequences of various species are shown in Table 2. For the species *Rhinomorinia sarcophagina* and *Gasterophilus haemorrhoidalis* the complete 18S rRNA gene could not be amplified from dry specimen.

Partial 16S rDNA sequences were obtained for ten species of the calyptrate superfamilies Hippoboscoidea, Oestroidea and Muscoidea. The sequence lengths vary from 476 to 480 bp, corresponding to the positions 707–1186 at the 3′ end of *Drosophila melanogaster* 16S rRNA gene (GenBank accession number X53506). When aligned with 9 additional dipteran 16S rDNAs retrieved from GenBank (Table 1), the resulting alignment is 494 positions

Table 1. List of species and accession numbers of genes analysed in this study

Suborder	Division	Section	Superfamily	Family	Species	Accession number	
						16S rDNA	18S rDNA
Nematocera	Tipulomorpha	–	Tipuloidea	Tipulidae	*Tipula* sp.	–	X89496*
	Culicomorpha	–	Culicoidea	Simuliidae	*Simulium damnosum*	AF081904*	–
Brachycera	Orthorrhapha	–	Tabanoidea	Tabanidae	*Chrysops niger*	–	AF073889*
		–	Asiloidea	Asilidae	*Laphria* sp.	–	U65157*
	Cyclorrhapha	Acalyptratae	Tephritoidea	Tephritidae	*Anastrepha fraterculus*	–	AF187101*
			Drosophiloidea	Drosophilidae	*Drosophila melanogaster*	–	M21017*
							M29800*
					Drosophila subobscura	AF126335*	–
		Calyptratae	Hippoboscoidea	Glossinidae	*Glossina m. morsitans*	AF072373*	–
					Glossina palpalis	–	AF322431
				Hippoboscidae	*Ornithomya avicularia*	–	AF322421
					Lipoptena cervi	AF322437	AF322426
					Ornithoica vicina	–	AF073888*
				Nycteribiidae	*Penicillidia* sp.	AF322435	AF322420
					Basilia sp.	–	AF322430
					Cyclopodia sp.	AF086864*	–
				Streblidae	*Brachytarsina* sp.	AF086865*	–
			Oestroidea	Calliphoridae	*Melinda viridicyanea*	–	AF322424
					Lucilia sp.	–	AF322425
					Lucilia cuprina	AF086858*	–
					Calliphora quadrimaculata	AF086857*	–
				Sarcophagidae	*Sarcophaga bullata*	AF322433	AF322419
				Tachinidae	*Nemoraea pellucida*	AF322432	AF322418
				Mystacinobiidae	*Mystacinobia zealandica*	AF086859*	–
				Rhinophoridae	*Rhinomorinia sarcophagina*	AF322434	–
				Gasterophilidae	*Gasterophilus haemorrhoidalis*	AF322439	–
				Hypodermatidae	*Oestromyia leporina*	–	AF322422
					Hypoderma diana	AF322438	AF322427
				Oestridae	*Cephenemyia stimulator*	AF322441	AF322429
				Cuterebridae	*Cuterebra* sp.	AF322440	AF322428
			Muscoidea	Muscidae	*Antipoda* sp.	AF086863*	–
					Musca sp.	AF322436	AF322423

*Sequences retrieved from GenBank.

long. The complete alignment contains three variable AT-rich regions corresponding to distal parts of stem-loop structures in *Drosophila melanogaster*-based secondary structure model (sequence accession number X53506; the secondary structure model obtained from www.rna.icmb.utexas.edu). Because these regions can not be aligned unequivocally, they are excluded from the phylogenetic analysis. After their exclusion, 409 positions are retained in the alignment, of which 121 variable positions provide eighty-one parsimony-informative sites. The distances among 16S rDNA sequences of calyptrate species (except for the identical sequences of two nycteribiids) varied from 2.6% to 25.5% (Table 2). The most diverged are the sequences of *Glossina*, *Gasterophilus* and the two nycteribiid species.

Phylogenetic analysis

In both Maximum Parsimony (MP) and Maximum Likelihood (ML) analyses, the 18S rDNA matrix, which includes twenty dipteran species, yields monophyletic Cyclorrhapha and Calyptratae, while the two acalyptrate species are arranged in a paraphyletic manner. This topology is robust and does not alter under a broad range of Ts/Tv ratios (1, 1.5, 2, 2.5 or 3; Fig. 3A). The interfamilial relationships within calyptrates remain largely unresolved. The only monophyletic groups retained under all parameters are the families Calliphoridae, Hypodermatidae, Nycteribiidae, and Hippoboscidae. The nodes retained in strict consensus are also the only ones supported by the bootstrap values higher than 50%. The hippoboscoid families never form a monophyletic clade and in majority of the trees, the nycteribiids branch as a sister group of *Hypoderma* + *Oestromyia*.

The 16S rDNA matrix including nineteen species reveals two different overall topologies. A setting of Ts/Tv ratio 1 and 1.5 results in fifteen trees with identical basal topologies, in which *Drosophila* branches as a sister group of all Calyptratae species. Within Calyptratae, strict consensus of the trees retains two separate lineages the hippoboscoid clade and the oestroid–muscoid clade (Fig. 3B). The hippoboscoid clade is one of the most robust elements in the tree supported by bootstrap value higher than 50%. An increase of Ts/Tv ratio to 2 and higher causes a complete flip of the topology moving *Gasterophilus* and *Hypoderma* to the base of the tree and places *Drosophila* deep into the crown

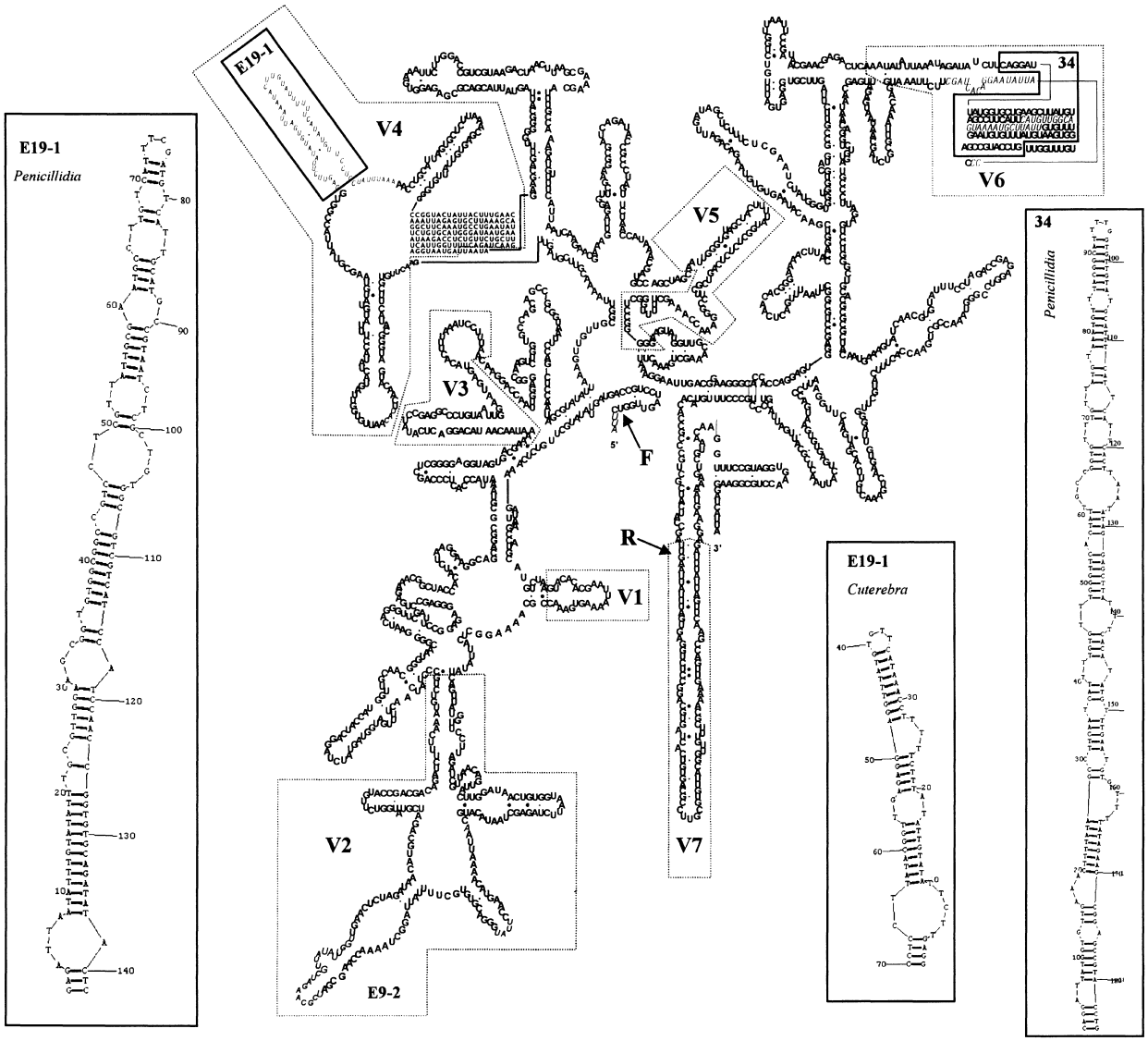

Figure 2. Secondary structure of *Drosophila melanogaster* 18S RNA (from Gutell Lab's Comparative RNA website; http://www.rna.icmb.utexas.edu). F, R = primers delimiting the region sequenced and analysed in this study. Stems and variable regions numbered according to Hancock *et al.* (1988). The positions excluded from the analysis printed in italics. The prediction of secondary structure of insertions determined in nycteribiids and *Cuterebra* are shown in boxes.

region as a sister group of Hippoboscoidea (not shown). In order to decide between these two competing topologies, the trees obtained under various parameters were used to calculate an empirical range of Ts/Tv ratio. The resulting values vary from 1.06 to 1.27 for 16S rRNA gene (and 1.6–1.9 for 18S rRNA gene). Regardless of the overall topology, the 16S rDNA sequences of muscoid species *Musca* sp. and *Antipoda* sp. cluster together with tachinid species *Nemoraea pellucida* and never form a separate monophyletic group.

Compared to the single-gene matrices, the combined matrix provides a better resolution of the lower-level phylogeny. In concordance with 18S rDNA results, the consensus of twelve trees obtained by MP analysis shows paraphyletic Acalyptratae and monophyletic Calyptratae. The calyptrates in concordance with 16S rDNA split into the hippoboscoid and oestroid lineages (Fig. 3C). Within hippoboscoid clade, the bat ectoparasites Nycteribiidae and Streblidae never form a monophyletic clade. Even if forced by constraint into one of the following monophyletic groups: Glossinidae + Nycteribiidae + Streblidae or Hippoboscidae + Nycteribiidae + Streblidae, the nycteribiids never branch as a sister group of streblids.

Secondary structures

Secondary structures are predicted for 18S rRNA variable regions corresponding to *Drosophila melanogaster* E9-2, E19-1, and 34 stem-loop structures (Fig. 2). The aims of

Table 2. Pairwise comparisons of nucleotide differences between dipteran species. The distances were calculated using the MEGALIGN program of the LASERGENE package (DNASTAR). Sequence distances: (A) 18S rDNA, (B) 16S rDNA

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A																					
1	Tipula	***																			
2	Chrysops	12.9	***																		
3	Laphria	7.7	6.7	***																	
4	Anastrepha	9.8	6.8	3.7	***																
5	Drosophila	9.4	6.9	4.3	4.7	***															
6	Melinda	9.9	7.4	3.8	4.7	3.5	***														
7	Lucilia	9.7	7.2	3.6	4.6	3.3	0.3	***													
8	Hypoderma	10.3	8.1	4.8	5.7	4.6	3.3	3.2	***												
9	Sarcophaga	9.6	6.8	4.3	4.3	3.5	2.1	1.9	3.5	***											
10	Cephenemyia	10.0	7.6	4.7	5.0	4.2	2.4	2.3	3.6	2.0	***										
11	Cuterebra	9.4	6.7	3.5	4.5	4.0	2.3	2.2	3.8	2.3	2.6	***									
12	Ornithomya	10.0	7.1	4.6	4.6	4.1	3.0	2.9	4.2	2.6	3.4	3.1	***								
13	Ornithoica	10.1	7.3	4.9	4.7	4.3	3.5	3.4	4.8	2.9	4.2	3.6	1.7	***							
14	Glossina	9.7	7.1	3.9	4.5	4.1	2.9	2.8	3.8	2.9	3.2	3.0	2.2	2.9	***						
15	Lipoptena	10.7	7.9	4.9	5.3	5.2	4.2	4.0	5.2	3.9	4.8	4.3	2.2	2.7	3.2	***					
16	Musca	9.6	6.6	3.9	4.5	3.6	2.5	2.3	3.4	2.5	2.8	2.8	2.8	3.4	2.8	3.6	***				
17	Oestromyia	10.6	8.4	4.9	6.2	5.2	3.8	3.7	0.5	4.0	4.1	4.2	4.7	5.3	4.1	5.6	3.7	***			
18	Nemoraea	9.4	7.4	3.8	4.3	3.3	1.8	1.7	3.1	1.9	2.2	2.1	2.7	3.4	2.7	3.9	2.1	3.7	***		
19	Penicillidia	14.1	10.8	5.5	4.8	5.1	4.5	4.3	5.2	4.0	4.6	4.3	4.3	4.5	4.6	5.0	4.2	5.5	4.2	***	
20	Basilia	15.8	12.0	6.9	6.2	6.3	5.4	5.3	6.2	5.1	5.7	5.4	5.4	5.6	5.6	6.1	5.0	6.8	5.2	4.1	***
B																					
1	Simulium	***																			
2	Drosophila	19.0	***																		
3	Lipoptena	22.9	14.3	***																	
4	Calliphora	19.8	11.3	16.8	***																
5	Lucilia	20.4	11.2	17.0	2.6	***															
6	Glossina	28.7	20.8	20.1	20.2	19.0	***														
7	Hypoderma	21.2	14.1	17.8	11.6	10.6	18.7	***													
8	Sarcophaga	21.3	12.8	17.0	5.4	6.1	21.1	13.5	***												
9	Cephenemyia	22.6	14.9	17.0	9.4	9.1	21.8	13.6	11.8	***											
10	Cuterebra	19.3	14.6	18.4	8.7	8.4	21.1	13.7	10.9	10.6	***										
11	Penicillidia	25.4	18.4	17.9	19.9	20.8	22.3	21.6	21.1	20.4	19.6	***									
12	Cyclopodia	25.4	18.4	17.9	19.9	20.8	22.3	21.6	21.1	20.4	19.6	0.0	***								
13	Brachytarsina	21.3	15.7	16.2	15.0	15.8	17.9	18.2	16.8	17.6	16.0	17.9	17.9	***							
14	Mystacinobia	20.4	9.5	15.5	9.2	8.4	22.2	12.4	8.9	13.6	12.4	17.9	17.9	16.0	***						
15	Gasterophilus	24.2	16.3	18.7	17.2	16.9	25.5	16.6	16.1	17.7	18.5	22.5	22.5	22.2	14.5	***					
16	Rhinomorinia	18.4	9.5	16.0	7.5	7.7	20.4	12.2	8.2	13.1	11.9	18.0	18.0	15.5	6.7	13.4	***				
17	Nemoraea	18.9	12.7	16.1	11.0	9.7	20.0	11.4	12.3	12.0	12.5	21.3	21.3	16.1	10.2	16.0	10.2	***			
18	Antipoda	22.1	11.5	17.9	10.9	9.4	22.4	14.7	12.6	12.3	13.6	17.7	17.7	14.7	9.1	15.7	8.9	9.5	***		
19	Musca	20.4	11.6	17.1	9.9	9.2	22.4	12.4	10.7	12.4	14.5	18.9	18.9	15.6	7.5	14.5	6.3	8.5	7.7	***	

*The family Oestridae presented in this tree includes subfamilies Oestrinae, Hypodermatinae and Gasterophilinae, which are treated as families in Table 1.

this analysis are to investigate the form of unique insertions in nycteribiids and *Cuterebra*, and to explore a possible phylogenetic significance of secondary-structure elements within the variable regions. The prediction of secondary structure for the insertions in V4 and V6 of nycteribiids reveals long unbranched stem-loop structures in the position where *Drosophila* and all other calyptrates have short conserved stem-loops E19-1 and 34 (Fig. 2). A similar but less pronounced stem-loop extension is present in V2 region of *Cuterebra* (Fig. 2).

Two of the stem-loop structures carry information potentially useful for phylogenetic inference. Within the V2 expansion segment, several different motifs are observed in the terminal region of E9-2 loop (Fig. 4b). While some of the motifs are unique – present in only single species – and could not therefore be used as a phylogenetic signal, two of them are more common and divide the taxa into two groups: GCAA motif shared by drosophilids, calliphorids and sarcophagids, and GAAA motif present in *Cephenemyia*, *Cuterebra*, *Hypoderma* and *Nemoraea*. In hippoboscoids, the terminal loop is biased towards Ts and similar bias was observed in Musca sp. A second feature derived from secondary structure is a presence of unpaired T within the helix of E19-1 loop, shared by both calliphorid species and Sarcophaga (Fig. 4A).

Discussion

Variations in rRNA genes

Of the two rRNA genes analysed in this study, a considerably higher degree of length variation is found among the 18S rDNA sequences. Among seven variable regions

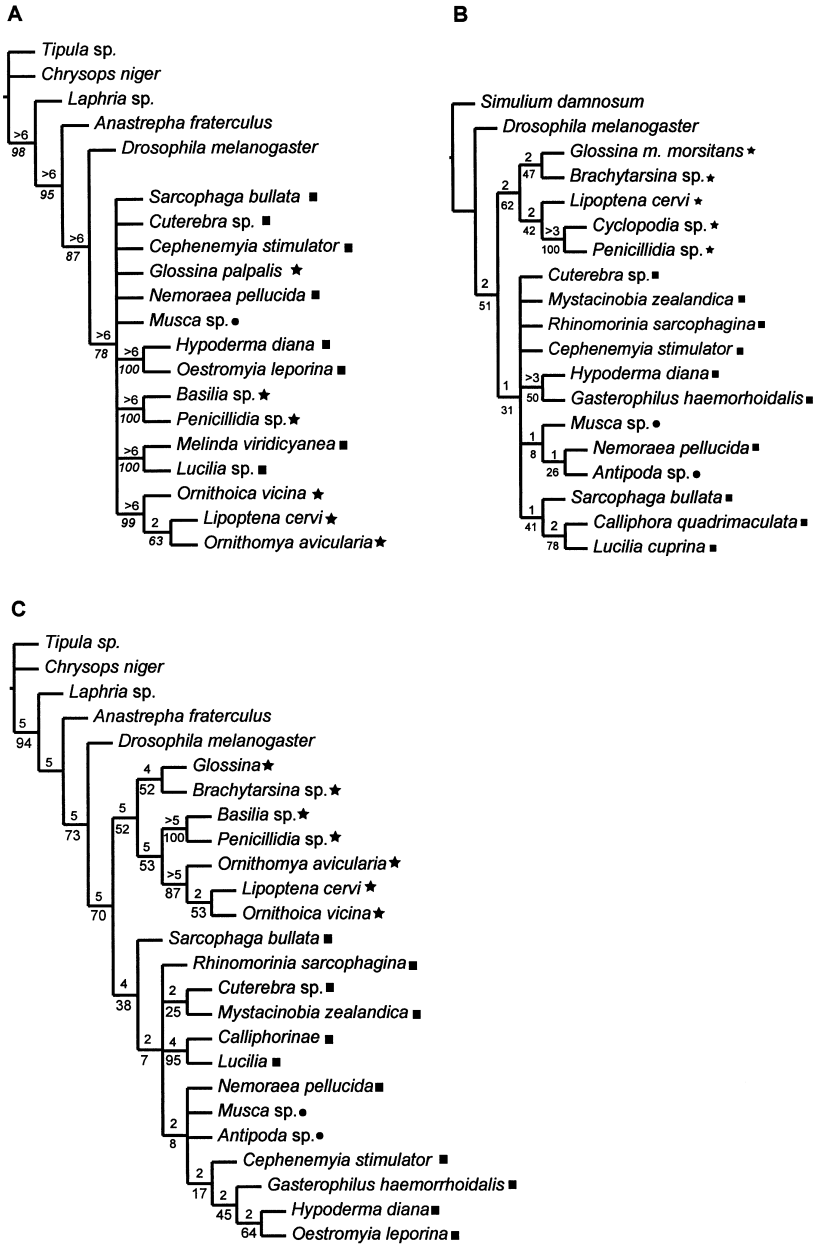

Figure 3. Consensus trees of maximum parsimony analyses of 18S rDNA, 16S rDNA, and combined matrices. (A) Strict consensus of thirty-six trees obtained by MP analyses of 18S rDNA with Ts/Tv ratios 1, 1.5, 2, 2.5, and 3. (B) Strict consensus of 16 trees obtained by MP analyses of 16S rDNA with Ts/Tv ratios 1 and 1.5. (C) Strict consensus of twelve trees obtained by MP analyses of combined DNA matrix with Ts/Tv ratios set to 1 for 16S rDNA and 1.7 for 18S rDNA. The numbers at nodes show the bootstrap values (below lines) and Bremer's indices (above lines). ■, Oestroidea; ★, Hippoboscoidea; ●, Muscoidea.

determined in 18S rRNA gene, the regions V4 and V6 are reportedly less conserved in insects when compared with vertebrate sequences (Baldridge & Fallon, 1991). Our data suggest that the dipteran 18S rRNA gene varies considerably in the V2 region, in addition to V4 and V6. The most striking is the presence of two long stem-loop insertions in the nycteribiid species located within V4 and V6 regions (Fig. 2). While the V6 insert is unique and is not present in other dipterans included in the data set, the location of V4 insert corresponds to the nematoceran insert occurring in the same region (represented by *Tipula*). A lack of any significant similarity with other brachyceran as well as nematoceran species indicates that whole DNA fragments were replaced and are in fact not homologous to corresponding sequence regions in other species.

Higher level phylogeny: the monophyly of Calyptratae

Analysis of the 18S rDNA sequence of twenty dipteran species demonstrates the ability of this gene to resolve the higher-level phylogeny within Diptera: a clear-cut separation is observed between Orthorrhapha and Cyclorrhapha, and also between Acalyptratae and Calyptratae (Fig. 3A). The paraphyletic arrangement of the two acalyptrate species supported by all 18S trees falls in line with the inference from the partial sequences of 28S rDNA (Vossbrinck & Friedman, 1989) that acalyptrates are a paraphyletic

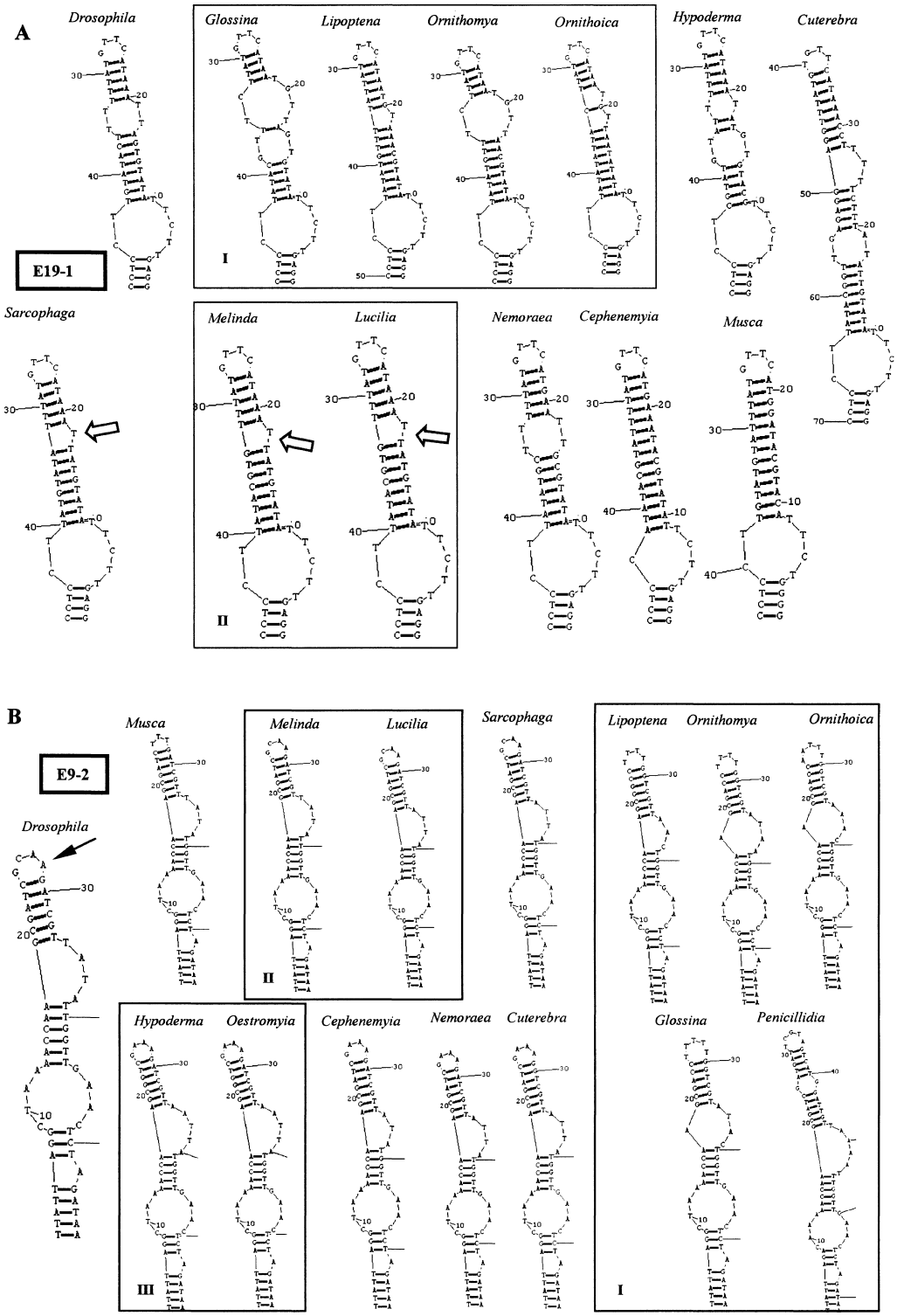

Figure 4. Comparative analysis of E19-1 (A) and E9-2 (B) stem-loop structures. Boxes: I, Hippoboscoidea; II, Calliphoridae; III, Hypodermatidae.

group with calyptrates branching closer to drosophilids than to tephritids. On morphological grounds, the paraphyly of acalyptrates had previously been proposed by Griffiths (1976). However, in his generally accepted classification, McAlpine (1989) treats calyptrates and acalyptrates as two monophyletic sister groups. He admits, though, that the acalyptrates monophyly was 'never satisfactorily established' and recognizes most of the characters used to support

the acalyptrates as 'plesiomorphic in relation to the ground-plan of Schizophora'. Although only two acalyptrate species are included in our analysis, their paraphyletic arrangement is well established in all trees and is always supported by high bootstrap and Bremer values (Fig. 3A).

Another well supported element of 18S and combined trees is a calyptrate monophyly (Fig. 3A,C). While interfamilial re-arrangements are observed among trees constructed by varying the Ts/Tv ratio, the Calyptratae monophyly is maintained in all trees. Although Calyptratae monophyly has never been seriously questioned on morphological grounds and is considered one of the best supported elements of dipteran classification (McAlpine, 1989), a tendency of drosophilids to cluster within Calyptratae has been reported previously in two molecular studies based on 28S rDNA (Vossbrinck & Friedman, 1989) and CO I and II (Bernasconi *et al*., 2000). In yet another molecular study, Carreno & Barta (1998) demonstrated a phylogenetic affinity of hippoboscid species *Ornithoica vicina* to *Drosophila melanogaster* and pointed out that such topology was in agreement with the concept of monophyletic Muscomorpha derived from morphological characters. However, because their data did not include any other higher Brachycera, they could not address the question of Calyptratae monophyly. In 16S rDNA analyses, the monophyly/paraphyly of Calyptratae in relation to *Drosophila* was dependent on Ts/Tv ratio used. In dipterans, the Ts/Tv ratio in the 28S rDNA is reported to vary from 0.8 in the loop regions to 2.02 in stem regions (Friedrich & Tautz, 1997) indicating that at least for this gene the assumption of any average ratio higher that 2 is unrealistic. To test whether the same applies to the genes used in our analyses, we calculated the empirical Ts/Tv ratios for the alternative topologies. The resulting ratio values varied from 1.6 to 1.9 for 18S, and from 1.06 to 1.27 for 16S rDNA. These values reflect the observation by Friedrich and Tautz in the 28S rDNA. The within-calyptrate position of *Drosophila* is therefore eliminated as a result of an overestimation of the Ts/Tv ratio. We consider the monophyly of Calyptratae well established and presume that the conflicting reports discussed above were most likely caused by usage of limited taxa sets (Bernasconi *et al*., 2000) or short DNA fragments (Vossbrinck & Friedman, 1989).

Hippoboscoidea

Several early morphological studies (Bequaert, 1954; Wenzel *et al*., 1966; Schlein, 1970) considered hippoboscids, on the one hand, and nycteribids + streblids, on the other, as two distant lineages. In his detailed anatomical study of thorax musculature, Schlein (1970) concluded that while Glossinidae and Hippoboscidae are undoubtedly sister taxa, the Nycteribiidae + Streblidae form a distant lineage which may not belong to Calyptratae, a view that was later supported by Pollock (1971).

In our study, two alternative positions of nycteribiids are observed, both of them placing this group clearly within Calyptratae. The 16S rDNA matrix recognizes the Hippoboscoidea superfamily as a monophyletic group resistant to broad range of parameters (Fig. 3B). Surprisingly, the bat ectoparasites Nycteribiidae and Streblidae never form a monophyletic clade, indicating that despite their similar bionomy, the two families may have evolved their association with bats independently. In contrast to a robust picture obtained from 16S rDNA matrix, the position of nycteribiids vary when 18S rDNA matrix is analysed (Fig. 3A). An application of empirical Ts/Tv ratio 1.7 (see above) in MP places the nycteribiids invariably as a sister group of *Hypoderma* + *Oestromyia*. The same topology is supported by distance-based analyses and all ML algorithms. The monophyly of Nycteribiidae + Hippoboscidae is only established when the Ts/Tv ratio is set to 3 and higher (MP) or when 'mean character difference' method is used as a distance measure method (Neighbour Joining; NJ). In a combined matrix, the 16S rDNA signal always prevails despite a lower number of parsimony-informative characters in 16S rDNA compared to 18S rDNA (81 vs. 146). This reveals a weak character of Nycteribiidae–Hypodermatidae association in 18S rDNA matrix. The instability of Nycteribiidae in the trees is obviously a long-branch phenomenon: the nycteribiids 18S rDNA sequences are the most diverged among Calyptratae (Table 2) with long insertions in the variable regions. We suppose that such replacements of whole DNA fragments within variable regions may be responsible for the loss of interfamilial phylogenetic information at the periphery of conservative and variable regions (Fig. 2).

Oestroidea

The overall topology of the calyptrate portion of the tree corresponds to the common view considering oestroids and muscoids more closely related to each other than they are to hippoboscoids (McAlpine, 1989; Yeates & Wiegmann, 1999; Fig. 3B,C). The interfamilial relationships within these lineages, although generally compatible with the morphology-derived classification, reveal several notable differences. In contrast to the morphology-based classification, no clear separation was observed between muscoid and oestroid species. While the 18S rDNA matrix does not provide sufficient information and places *Musca* within a polytomy together with oestroid species, the 16S rDNA sequences of muscoid species cluster invariably together with tachinid species *Nemoraea pellucida* and are never found to form a sister group of oestroids. Moreover, they always branched deep inside the oestroid cluster and do not exceed the average length of branches in calyptrates suggesting that their position within oestroid in 16S rDNA matrix is stable and is unlikely to be due to a long branch attraction or any other algorithm artifact. This observation seems to complement the results of Bernasconi *et al*.

(2000) who failed to find separate oestroid and muscoid lineages using COI and COII sequences. In their study, the oestroid family Calliphoridae branched within muscoid group represented by Scathophagidae, Muscidae, Fanniidae, and Anthomyiidae. The comparison of their results and the trees presented here leads to a conclusion that both superfamilies may represent intertangled polyphyletic assemblages of calyptrate families.

The relationship between Tachinidae and Sarcophagidae differs from the most recent morphology-based classification. The monophyly of Tachinidae + Sarcophagidae was first suggested by Pape (1992). Later, Rognes (1997) showed that the presumed synapomorphy of this group (fused condition of dorsolateral processes of the aedeagus) becomes symplesiomorphic after inclusion of *Mystacinobia* and hence can not be used to validate the Tachinidae + Sarcophagidae clade. Although his own morphological analysis also produced the monophyly of Tachinidae + Sarcophagidae, it did not rest on any unique synapomorphy: in fact, two out of four characters supporting this node, were polymorphic in several families, including Sarcophagidae (12, 22 in Rognes, 1997), and yet another character was admitted to display exceptions, especially in Tachinidae. In our analysis, Tachinidae and Sarcophagidae never occur as sister groups. In 18S rDNA analysis the position of *Sarcophaga* is unstable, while in all other analyses, 16S rDNA and combined, this species branches in the vicinity of calliphorids, either in a monophyletic (16S; Fig. 3B) or paraphyletic (combined; Fig. 3C) manner (Fig. 3A,B). This topology is resistant to inclusion of Rogne's morphological data even if weighted 20-fold higher than the molecular part of matrix. Although in conflict with the latest morphological studies, this phylogenetic position of sarcophagids is not novel and agrees with an older opinion of several authors (see McAlpine, 1989). In order to investigate the possible relationships among sarcophagids, tachinids and calliphorids more closely, we analysed secondary structure features of variable loops. Such an approach previously has been shown to provide relevant phylogenetic information in crustaceans (Crease & Taylor, 1998). Because the parts of sequences that we used for this study were excluded from the phylogenetic analysis presented above, they provide additional information, independent of the alignments. Out of four excluded variable regions of 18S rDNA, two fragments yield secondary structures displaying phylogenetically useful traits. In both cases, these traits are identical for *Sarcophaga* + both calliphorid species. Within V2 expansion segment, several different motifs are present in the terminal portion of E9-2 loop (Fig. 4B; see results). The distribution of these features thus favours the McApine's view grouping Sarcophagidae + Calliphoridae and Tachinidae + Oestridae.

Several practices have been employed in the classification of the four obligatory vertebrate-associated endoparasitic groups. These flies were treated either as separate families (Hennig, 1973) or as subfamilies within the family Oestridae (McAlpine, 1989). Alternatively, Pont (1980) recognized two families, Oestridae encompassing Hypodermatinae + Oestrinae, and the separate family Gasterophilidae. While the monophyly of Hypodermatidae + Oestridae + Gasterophilidae is supported in our analysis by combined matrix, the position of Cuterebridae is not clear. A possible relationship of this species to other oestrids was implied by the terminal motif of E9-2 structure (Fig. 4B). Such a position, however, is not supported by phylogenetic analyses. In single gene matrices, the position of *Cuterebra* is unstable: in the consensus tree; it branches as a separate lineage within a polytomy encompassing all calyptrate species (Fig. 3A,B). The combined matrix clusters the *Cuterebra* together with *Mystacinobia*, rather than oestrids (Fig. 3C). A conflict is observed between phylogenetic signals derived from the matrices and the features of secondary structure in *Cuterebra* and *Musca*. However, the overall instability of interfamilial relationships within oestroids, as well as extremely short branches provided by ML analyses for interfamilial level, indicate that the position of a few families may change after more data are obtained and analysed, and better fit may be achieved between secondary structure-derived characters and matrices-based phylogeny. All analyses, regardless of the character exclusion, Ts/Tv ratio and method employed, support unequivocally the calyptrate origin of Mystacinobiidae and their relationships to other oestroid families. This is in agreement with recent morphological and molecular studies (Rognes, 1997; Gleeson *et al.*, 2000).

Apart from the monophyly and mutual relationships of endoparasitic vertebrate-associated families (Hypodermatidae, Gasterophilidae, Oestridae, Cuterebidae), the question of particular importance from the evolutionary point of view is their origin within oestroid clade, i.e. whether ancestors of these endoparasites were free living or facultatively parasitic. Traditionally, Calliphoridae, a large family with many facultative and obligatory endoparasitic species, has been considered as a possible evolutionary key family for obligatory parasitic lineages within Oestroidea (McAlpine, 1989; Rognes, 1997). In our study, however, both calliphorid species examined did not seem to share an ancestor with either of the three vertebrate-associated oestroid families (Hypodermatidae, Gasterophilidae, and Oestridae), which together form a monophyletic clade. Instead, tachinids rather than calliphorids, are closer to those three oestroid families in all combined analyses. Because calliphorids might have paraphyletic character (Rognes, 1997) and we included only the sequences of two calliphorid subfamily members (Calliphorinae and Lucillinae) we conclude that tracing down the origin of vertebrate-associated obligatory parasites, and finding their closest extant relatives, will require inclusion of other calliphorid lineages.

Experimental procedures

Species analysed

The specimens examined in this study are listed in Table 1. In order to avoid confusion, the subfamilies Cuterebrinae, Gasterophilinae, Hypodermatinae, and Oestrinae according to McAlpine (1989) are in this text referred to as families.

DNA extraction, PCR amplification and sequencing

Total DNA was extracted from fresh, frozen, dry or ethanol preserved specimens. Individual flies or tissue portions were grinded in liquid nitrogen after the addition of 10 mM Tris (pH 8.0), 0.1 M EDTA and 0.5% SDS. Proteinase K (100 μg/ml) was subsequently added and the mixture was incubated at 50 °C for 3 h. The mixture was cooled to room temperature and equal amount of 1:1 phenol:chloroform was added, mixed and centrifuged. The aqueous phase was either mixed with 2.5 volumes of buffer QG and passed through a spin column (QIAquick gel extraction kit) to purify the extracted DNA or an equal volume of 100% isopropanol was added to precipitate the DNA, which was stored at −20 °C until required. Alternatively, a small portion of tissue from a fly was first crushed in a plastic homogenizer to which 200 μl of 5% Chelex 100 resin (Bio-Rad) in water was added and incubated at 50 °C for 3 h for tissue degradation. The mixture was subsequently heated at 95–100 °C for 15 min and stored at −20 °C until required. Extracts were centrifuged at 5000 ***g*** for 5 min and an aliquot of the supernatant was used for PCR (Edwards & Hoy, 1993). These DNA preparations were stable at least for 2 years at −20 °C.

Primers 18e and 18g (Hillis & Dixon, 1991) were used to amplify the 1.8 kb long fragment of 18S rRNA gene, and primers mt33 and mt34 (insect mt-DNA set, Biotechnology Laboratory, University of British Columbia; http://www.biotech.ubc.ca/services/naps/primers.html) were used to amplify 500 bp long fragment of 16S rRNA gene. Amplifications were performed in a 50-μl reaction mixture containing 1 × PCR buffer (Top-Bio), 200 μM dNTP, 2 mM $MgCl_2$, 2.5 units Taq polymerase (Top-Bio), and 20 pmol of each primer. Temperature cycling generally consisted of a 2-min initial denaturation at 94 °C followed by 30 cycles each including 94 °C for 30 s, 53 °C for 30 s, and 72 °C for 1 min and 30 s, and a final extension at 72 °C for 7 min. Alternatively, a step-up temperature cycling was performed with a 2-min initial denaturation at 94 °C, 10 cycles each including 94 °C for 30 s, 45 °C for 30 s, and 72 °C for 1 min and 30 s, followed by 15 cycles each including 94 °C for 30 s, 53 °C for 30 s, and 72 °C for 1 min and 30 s and the final extension at 72 °C for 7 min. PCR reactions were performed in either Perkin-Elmer Gene amp PCR system 2400 or MJ Research PTC 200 thermal cycler.

PCR products were purified using a gel extraction kit (Qiagen) and used either directly for sequencing or ligated into the pGemT-easy vector (Promega), and sequenced using M13 reverse and forward primers in an ABI PRISM sequencer (Perkin–Elmer model 310). To sequence a complete 18S rRNA gene, we used internal primers 18h, 18k, and 18i according to Hillis & Dixon (1991), and designed two additional primers, 18x: 5′ GCG TCA GAG GTG AAA TTC TT 3′ and 18y: 5′ GGT CTA AGA ATT TCA CCT CT 3′. The position of 18x and 18y primers according to the *Drosophila melanogaster* 18S rRNA gene (GenBank accession number M21017) corresponds to the positions 980–999 and 985–1004, respectively. Sequence was determined for both strands of either the PCR product or the respective clone. The obtained sequences were deposited in GenBank and accession numbers are provided in Table 1.

Alignments and phylogenetic analysis

Three different nucleotide sequence alignments were generated. (1) 16S rDNA alignment containing 19 dipteran species (17 families); with *Simulium damnosum* used as an outgroup. (2) 18S rDNA alignment containing twenty species (fifteen families) with one nematoceran species, *Tipula* sp., and two orthorrhaphan species, *Chrysops niger* and *Laphria* sp. as outgroups, and (3) a combined alignment representing twenty-five species of sixteen families. In order to create this alignment, the 18S and 16S rDNA matrices were concatenated. The combined matrix contained all calyptratae taxa for which at least one gene (18S or 16S rDNA) was available (except for 16S rDNA of *Cyclopodia* sp. which was identical to that of *Penicillidia* sp.). The sequences in this matrix were combined as follows: (i) for seven taxa, the 16S and 18S rDNA sequences obtained from single species were combined (listed as individual species in the combined tree); (ii) because only genera were determined for both rDNAs of *Penicillidia* and 18S rDNA of *Lucillia*, the combination of 16S and 18S rDNA sequences are considered to represent these genera rather than particular species; (iii) the 16S rDNA of *Calliphora quadrimaculata* and 18S rDNA of *Melinda viridicyanea* were combined to represent the subfamily Calliphorinae (Combined matrix).

Sequence alignments were produced using CLUSTAL algorithm as implemented by the MEGALIGN program of the LASERGENE package (DNASTAR Inc.) and corrected according to the model of *Drosophila melanogaster* 18S rRNA secondary structure retrieved from htttp://www.rna.icmbutexas.edu. The variable regions, which could not be aligned reliably, were removed from the phylogenetic analyses (Fig. 2). Both 18S and 16S rDNA alignments are available upon request from the authors. Multiple alignments were analysed using PAUP* 4.0b4 (Swofford, 1998). Maximum parsimony analyses were performed by TBR algorithm with fifty randomizations of sequence order. For maximum likelihood analysis, several models and parameters settings were tested (HKY, F84, JC, Kimura-2 parameters model, Ts/Tv ratio 1, 1.5, 2, 2.5, and 3 with five randomizations of sequence order). In all analyses, gaps were treated as missing data. The bootstrap support was calculated from 1000 replications; Bremer (Decay) indices were determined by releasing constraint on tree length (the PAUP* command KEEP = length) till complete polytomy was obtained. The empirical value of Ts/Tv ratio for each topology was calculated from a list of character changes produced by a PAUP* command DESCRIBETREES/CHGLIST = YES.

Reconstruction of secondary structures was based on a putative structure of 18S rRNA gene of *Drosophila melanogaster* obtained from the Gutell Lab's Comparative RNA website (http://www.rna.icmb.utexas.edu). Using this model, the proximal ends of a given loop were determined in 18S rDNA sequence alignment and the corresponding regions in other sequences were folded by RNA structure v.3.5 retrieved from the website of Turner group RNA biophysical chemistry (http://rna.chem.rochester.edu/index.html).

Acknowledgements

We thank Jan Minár (Institute of Entomology, Academy of Sciences of the Czech Republic) for providing the samples

of *Gasterophilus haemorrhoidalis* and *Cephenemyia stimulator*. Dry samples of *Nemoraea pellucida*, *Melinda viridicyanea*, *Rhinomorinia sarcophaginia* and *Hypoderma diana* were kindly provided by Prof. Rudolf Rozkošný (Masaryk University, Brno); the specimen of *Cuterebra* sp. was provided by John Stireman (University of Arizona, Tucson). We thank Andrey Tatarenkov (University of California, Irvine) for critical reading of the manuscript. This work was supported by grant no. A6022801 (Grant Agency of the Academy of Sciences, Czech Republic) and by grant nos 96066 and MSM 123100003 (Ministry of Education, Czech Republic).

References

Baldridge, G.D. and Fallon, A.M. (1991) Nucleotide sequence of a mosquito 18S ribosomal-RNA gene. *Biochem Biophys Acta* **1089**: 396–400.

Bequaert, J.C. (1954–7) The Hippoboscidae or louse-flies (Diptera) of mammals and birds. Part II. *Entomologica Am* **34–36**: 1–611.

Bernasconi, M.V., Valsangiacomo, C., Piffaretti, J.-C. and Ward, P.I. (2000) Phylogenetic relationships among Muscoidea (Diptera: Calyptratae) based on mitochondrial DNA sequences. *Insect Mol Biol* **9**: 67–74.

Carreno, R.A. and Barta, J.R. (1998) Small subunit ribosomal RNA genes of tabanids and hippoboscids (Diptera: Brachycera): evolutionary relationships and comparison with other Diptera. *J Med Entomol* **35**: 1002–1006.

Crease, T.J. and Taylor, D.J. (1998) The origin and evolution of variable-region helices in V4 and V7 of the small-subunit ribosomal RNA of brachiopod crustaceans. *Mol Biol Evol* **15**: 1430–1446.

Edwards, O.R. and Hoy, M.A. (1993) Polymorphism in 2 Parasitoids detected using random amplified polymorphic DNA-polymerase chain reaction. *Biol Control* **3**: 243–257.

Friedrich, M. and Tautz, D. (1997) An episodic change of rDNA nucleotide substitution rate has occurred during the emergence of the insect order Diptera. *Mol Biol Evol* **14**: 644–653.

Gleeson, D.M., Howitt, R.L.J. and Newcomb, R.D. (2000) The phylogenetic position of the New Zealand batfly, Mystacinobia zelandica (Mystacinobiidae; Oestroidea) inferred from mitochondrial 16S ribosomal DNA sequence data. *J Roy Soc New Zeal* **30**: 155–168.

Griffiths, G.C.D. (1976) Comments on some recent studies of tsetse-fly phylogeny and structure. *Syst Entomol* **1**: 15–18.

Hagan, H.R. (1951) *Embryology of the Viviparous Insects*. Roland Press, New York.

Hancock, J.M., Tautz, D. and Dover, G.A. (1988) Evolution of secondary structure and compensatory mutations of the ribosomal RNAs of *Drosophila*. *Mol Biol Evol* **5**: 393–414.

Hennig, W. (1973) Diptera (Zweiflüger). *Handb Zool Berlin* **4**: 1–200.

Hillis, D.M. and Dixon, M.T. (1991) Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quart Rev Biol* **66**: 410–453.

McAlpine, J.F. (1989) Phylogeny and classification of the Muscomorpha. In: *Manual of Nearctic Diptera* (McAlpine, J.F., ed.), Monograph 32, Vol. 3, pp. 1397–1518. Research Branch, Agriculture Canada.

Meier, R., Kotrba, M. and Ferrar, P. (1999) Ovoviviparity and viviparity in diptera. *Biol Rev* **74**: 199–258.

Pape, T. (1992) Phylogeny of the Tachinidae family-group (Diptera: Calyptratae). *Tijds Entomol* **135**: 43–86.

Pollock, J.N. (1971) Origin of tsetse flies: a new theory. *J Entomol* **40**: 101–109.

Pont, A.C. (1980) 94. Family Gasterophilidae, 95. Family Oestridae. In: *Catalogue of the Diptera of the Afrotropical Region* (Crosskey, R.W., ed.), pp. 883–888. British Museum, Natural History, London.

Rognes, K. (1997) The Calliphoridae (Blowflies) (Diptera: Oestroidea) are not a Monophyletic Group. *Cladistics* **13**: 27–66.

Schlein, Y. (1970) A comparative study of the thoracic skeleton and musculature of Pupipara and the Glossinidae (Diptera). *Parasitology* **60**: 327–373.

Swofford, D.L. (1998) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, v.4. Sinauer Associates, Sunderland, MA.

Vossbrinck, C.R. and Friedman, S. (1989) A 28s ribosomal RNA phylogeny of certain cyclorrhaphous Diptera based upon a hypervariable region. *Syst Entomol* **14**: 417–431.

Wenzel, R.L., Tipton, V.J. and Kiewlicz, A. (1966) The streblid batflies of Panama. In: *Ectoparasites of Panama*. Field Museum of Natural History, Chicago, IL.

Yeates, D.K. and Wiegmann, B.M. (1999) Congruence and controversy: towards a higher-level phylogeny of Diptera. *Ann Rev Entomol* **44**: 397–428.