

Molecular sequence data resolves basal divergences within Simuliidae (Diptera)

JOHN K. MOULTON

Department of Entomology, University of Arizona, Tucson, Arizona, U.S.A.

Abstract. Basal divergences within Simuliidae are inferred from portions of the large nuclear ribosomal subunit (28S), elongation factor-one alpha (EF-1 α), dopa decarboxylase (DDC), phosphoenolpyruvate carboxykinase (PEPCK), the small mitochondrial ribosomal subunit (12S) and subunit two of NADH dehydrogenase (ND2). Sequences from 28S and EF-1 α are presented for a thaumaleid, a chironomid and twenty-five of the thirty-two currently recognized simuliid genera. For DDC, PEPCK, 12S and ND2 the taxon sampling is less extensive, but includes both aforementioned outgroup taxa, *Parasimulium* Malloch, *Simulium* Latreille and genera representing each of the two to three major suprageneric taxa of various authors. Phylogenetic analyses of 28S, EF-1 α , PEPCK and DDC sequences yield strong support for the basal position of *Parasimulium* (Parasimuliinae) and division of the remaining simuliids (Simuliinae) into monophyletic sister taxa, Prosimuliini and Simuliini. These groupings are entirely concordant with those recovered in the most intensive analysis of morphological data.

Introduction

Approximately 1660 species of black flies representing no fewer than thirty-two genera have been described, with Antarctica being the only major land mass on which they are absent (Crosskey, 1988; Crosskey & Howard, 1997). They are a homogeneous group, particularly morphologically, in spite of their antiquity. Based predominantly upon geographical distributions of purportedly closely related taxa, Rubtzov (1956) and Dumbleton (1963) hypothesized the major lineages within Simuliidae were already in place by the middle of the Mesozoic. Two Jurassic-age fossils, *Simulidium priscum* Westwood and *Pseudosimulium humidum* Westwood, that corroborated their views at the time have since been relegated to other families by Rohdendorf (1964) and Craig (1977), respectively.

Since these reassignments, several additional Mesozoic-age fossil simuliids have been described. *Simulimima grandis* Kalugina (Kalugina & Kovalev, 1985), known only from a single pupal depression dating back to the middle Jurassic, \approx 160–170 mybp, was assigned to Simuliidae by Crosskey (1991), who likened it to *Prosimulium* Roubaud. Jell & Duncan (1986) reported several simuliid fossils, e.g. an adult, dissociated wing and fourteen larval depressions

from the Koonwarra fossil beds of southern Victoria, Australia, which date back to the Aptian phase of the lower Cretaceous, \approx 113–119 mybp. Currie (1988) considered only the larval depressions to be simuliids and noted similarities between them and an extant Australian taxon. Kalugina (1991) described three simuliids, *Baisomyia incognita*, *Gydarina karabonica* and *Kovalevimyia lacrimosa*, from fossil adults preserved in Upper Jurassic or Lower Cretaceous rocks from Transbaikalia and affiliated them with *Gymnopais* Stone/Twinnia Stone & Jamnback, *Gymnopais* and *Prosimulium*, respectively. The fact that all of these Mesozoic-age fossil simuliids can arguably be assigned to extant genera serves as a testament to the antiquity and morphological homogeneity of the group.

This large degree of morphological uniformity within Simuliidae has hampered attempts to infer phylogenetic relationships among its constituent genera. Many of the characters useful in elucidating basal lineages are simple in nature and differ primarily by gradation (Dumbleton, 1963; Crosskey, 1969). Different views of the delineation of these characters into states and their relative importance as phylogenetic indicators has had a profound impact on the limits of basal clades within Simuliidae. As a result, more than ten classification schemes have been proposed for Simuliidae since the turn of the twentieth century (Edwards, 1934; Enderlein, 1937; Smart, 1945; Rubtzov, 1956, 1974; Shewell, 1958; Grenier & Rageau, 1960; Stone, 1965; Crosskey, 1969, 1981, 1988; Currie, 1988; Py-Daniel, 1990;

Correspondence: John K. Moulton, Department of Entomology, Forbes Building, Rm. 410, University of Arizona, Tucson, AZ 85721, U.S.A. E-mail: kmoulton@ag.arizona.edu

Yankovsky, 1992; Py-Daniel & Moreira Sampaio, 1994). Wygodzinsky & Coscarón (1973) were the first to view simuliid relationships from a strictly phylogenetic standpoint, and Currie (1988) and Py-Daniel (1990) were the first to infer basal relationships within Simuliidae using cladistic methods.

It is difficult to make direct comparisons between or formulate a general consensus from the aforementioned classifications because most consider only a limited fauna. Nonetheless, a sister group relationship between *Parasimulium* Malloch and the remaining Simuliidae is clearly one of the most widely accepted features, although there is opposition even at this basic level. Workers following this common scheme subsequently divide the remainder of Simuliidae into two or three major groups. The arrangement and composition of these groups, however, often differ substantially, thereby implying considerable differences of opinion about phylogenetic relationships and the relative importance of characters used to infer them.

Molecular studies on black flies have not provided much insight into resolving basal relationships, as they have primarily been limited to the separation of closely related species and elucidation of relationships among them (Post & Flook, 1992; Brockhouse *et al.*, 1993; Xiong & Kocher, 1993; Tang *et al.*, 1995, 1996). Xiong & Kocher (1991), Pruess *et al.* (1992) and Tang *et al.* (1995) dealt somewhat with simuliid basal relationships but limitations in quantity of data (<350 characters) and taxon sampling (<5 simuliid genera) preclude the inference of meaningful conclusions.

As a result of the limited amount and confusing nature of morphological variation and the paucity of other sources of information, such as molecular or cytological, even the most basic taxonomic divisions within Simuliidae are subject to debate. The large number of conflicting phylogenies and classification schemes based upon morphological evidence indicates the need to examine alternative sources of phylogenetic information. This research is part of a large undertaking to infer relationships within Simuliidae using molecular sequence data. Partial nucleotide sequences from four nuclear genes, 28S, EF-1 α , DDC and PEPCK comprising 1385–1507, 1239, 519 and 543 characters, respectively, and two mitochondrial genes, 12S and ND2 comprising 514 (only 294 for *Austrosimulium bancrofti* Taylor partial sequence from GenBank) and 454 characters, respectively, are presented herein. Large subunit nuclear rDNA and EF-1 α sequences are presented from a chironomid, a thaumaleid and simuliids representing all but seven of the known genera and all major groups ever hypothesized. For DDC, PEPCK and 12S, smaller though still representative numbers of simuliid genera were sequenced in addition to the aforementioned outgroups. Taxon sampling for ND2 is the least intensive. These sequences (and amino acid sequences of DDC, PEPCK and ND2) were subjected to phylogenetic analyses using parsimony, minimum evolution and neighbour-joining criteria, and the resulting inferences used as a foundation against which previous systematic research on Simuliidae is reviewed.

Materials and methods

Genomic sampling

The simuliid fossil record, although sparse, convincingly dates the family back to at least the middle Jurassic. Since the ultimate goal of this research is inference of relationships throughout the family, genes were selected in the hope that they would provide information useful for reconstructing basal, intermediate and shallow divergences. The nuclear large ribosomal subunit has been shown to be useful for reconstructing both deep (Perasso & Baroin Tourancheau, 1993; Hoc *et al.*, 1993) and shallow (Vossbrinck, 1989; Pélandakis & Solignac 1993) divergences. EF-1 α , PEPCK and DDC are among fourteen nuclear protein-coding genes considered by Friedlander *et al.* (1992) to be promising candidates for elucidating higher-level phylogenetic relationships. In follow-up studies, Cho *et al.* (1995), Friedlander *et al.* (1996), Fang *et al.* (1997) and Mitchell *et al.* (1997) corroborated these exploratory studies through concordance with well established relationships within Lepidoptera. The divergence times within Lepidoptera for which these genes were phylogenetically informative were hypothesized to coincide with many within Simuliidae.

The 12S rRNA and ND2 genes were selected largely because mitochondrial loci were desired for the sake of completeness and because 16S and more slowly evolving mitochondrial protein coding genes, particularly ND4, exhibit little useful variation among simuliids or between mosquitoes and simuliids (Xiong & Kocher, 1991; J. K. Moulton, unpublished observations).

Taxon sampling

The choice of outgroup was influenced by morphological data (Hennig, 1973; Wood & Borkent, 1989; Oosterbroek & Courtney, 1995), molecular data (Pawlowski *et al.*, 1996) and, to a lesser degree, availability of published sequences. Inclusion of a thaumaleid in all data sets is deemed especially important in light of the findings of Pawlowski *et al.* (1996). Because monophyly of Culicomorpha has been demonstrated both morphologically (Hennig, 1973; Wood & Borkent, 1989; Oosterbroek & Courtney, 1995) and molecularly (Friedrich & Tautz, 1997), and cursory analyses of partial data matrices in the early phases of this study showed concordance with respect to the rooting of Simuliidae, the minimal number of outgroup taxa included here seemed adequate. Some outgroup sequences were obtained from the GenBank database (Table 1).

Generic concepts in Crosskey's (1988) conspectus of Simuliidae of the world served as a reference for the selection of simuliid taxa, with some modification in light of the findings of Currie (1988). Generic placement of Australian species not belonging to *Austrosimulium* Tonnoir or *Simulium* Latreille, e.g. the Australian '*Cnephia*' or prosimuliines of authors, is problematical. I treat these in a manner similar to that of Crosskey (1988) and Currie (1988), i.e. assigned to *Cnephia* Enderlein but obviously belonging elsewhere. *Austrosimulium*

colboi Davies & Györkös is also erroneously assigned (J. K. Moulton, unpublished observations). Quotation marks, e.g. '*Cnephia*' *strenua* Mackerras & Mackerras, are used throughout to stress the inappropriateness of these generic assignments.

Taxon sampling for 28S and EF-1 α (Table 1) was limited by the lack of suitably preserved material of seven simuliid genera and the cost of including more than a single representative from more than a few key genera. More than one species of *Austrosimulium*, *Gigantodax* Enderlein, *Greniera* Doby & David and *Simulium* were included in the 28S and EF-1 α data sets in an attempt to shorten potentially long branches, test the monophyly of *Austrosimulium* and *Greniera* and determine relationships within *Simulium* s.l. Although the latter points are of no concern here, all taxa were included in the 28S and EF-1 α analyses performed herein. Higher level divergences within black flies will be reported elsewhere (J. K. Moulton, unpublished observations). The difference in taxon sampling between 28S and EF-1 α is attributable to an emphasis placed upon the former gene after it was apparent that variation in the latter was at or near saturation.

Preliminary results from phylogenetic analyses of the former two genes, aspects of various classification schemes (Rubtzov, 1974; Crosskey, 1981, 1988; Currie, 1988) and statements in the literature (Crosskey, 1969; Currie, 1988) about the degree of 'primitiveness' of some genera influenced the selection of taxa for PEPCK and DDC. Differences in the taxa sampled for these two genes are attributable to differential success in amplification. Based upon morphological features, the substitution of '*Cnephia*' *strenua* for *Paracnephia* Rubtzov and *Simulium congareanarum* Dyar & Shannon for *Simulium enciso* Vargas & Díaz Nájera is probably justifiable. The sparse taxon sampling for 12S and ND2 is attributable to a lack of suitably preserved material at the time these loci were considered and the speed with which the author became dissatisfied with their phylogenetic informativeness.

Araucnephia Wygodzinsky & Coscarón, *Araucnephioidea* Wygodzinsky & Coscarón, *Cnesiamima* Wygodzinsky & Coscarón, *Lutzsimulium* d'Andretta & d'Andretta (including *Kempfsimulium* Py-Daniel & Nunes de Mello), *Tlalocomyia* Wygodzinsky & Díaz Nájera, *Sulcicnephia* Rubtzov & *Levitinia* could not be included in this study owing to the lack of suitably preserved material. Based upon morphological evidence, *Araucnephia* is related to *Araucnephioidea* (Wygodzinsky & Coscarón, 1973); *Cnesiamima* is related to *Paraustrosimulium* Wygodzinsky & Coscarón (Wygodzinsky & Coscarón, 1973; Coscarón, 1985); *Tlalocomyia* is related to *Mayacnephia* Wygodzinsky & Coscarón (Wygodzinsky *et al.*, 1970; Wygodzinsky & Coscarón, 1973); *Levitinia* is related to *Gymnopais* and *Twinnia* (Currie, 1988). Although the true affinities of *Sulcicnephia* are unknown, the well formed cocoon, deep pedisulcus and well sclerotized, apically armed parameres suggest a position near *Simulium* s.l. Therefore, the absence of *Araucnephia* or *Araucnephioidea* may represent the only significant gap in the simuliid sampling other than for *Parasimulium* (*Astoneomyia*) Peterson. No attempts were made to include *Piezosimulium jeanninae* Peterson because the three frag-

mentary pharate adults upon which the description is based (Peterson, 1989) are now believed to be mermithiforms of *Prosimulium neomacropyga* Peterson or a closely related species (Crosskey & Howard, 1997).

DNA extraction

Total DNA was extracted from a single fresh, frozen or absolute ethanol-preserved individual of each species. Larval head capsules, pupal gills, or wings, legs and terminalia of adults were saved as vouchers and are deposited in the University of Arizona Arthropod Collection. Whole specimens, minus the pieces removed as vouchers, were homogenized in 750 μ l of lysis buffer (50 mM Tris, pH 8.0; 50 mM EDTA, pH 8.0; 2% SDS, 75 mM NaCl; 50 mM sucrose) plus 10 μ l of 20 mg/ μ l proteinase K. The homogenates were incubated at 55°C for 12–24 h. Samples were extracted twice with phenol:chloroform:isoamyl alcohol (Sigma) (25:24:1) and once with chloroform:isoamyl alcohol (24:1). One-tenth volume of 3 M NaOAc and 1.7 \times volume of –20°C absolute ethanol were added to precipitate the DNA, which was then pelleted in a microcentrifuge. DNA was washed with one milliliter of 70% and 95% EtOH, air-dried for at least 6 h, and resuspended overnight at room temperature in 100 μ l of TE with RNase A. Samples were stored at –20°C.

DNA amplification

Amplifications were performed in a 50- μ l solution containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 3.5–4.0 mM MgCl₂, 10 pmol of each primer, one unit of Taq polymerase (Life Technologies, GIBCO BRL, Gaithersburg, Maryland), 0.2 mg of T4 gene 32 protein (Ambion Incorporated, Austin, Texas) and 5–10 ng of template DNA. Primers used in this study are listed in Table 3. Cycling parameters for 28S were a 5-min soak at 94°C and 37 cycles of 94°C for 30 s, 64°C for 40 s and 72°C for 2.5 min. Cycling parameters for EF-1 α amplification were a 5-min soak at 94°C and 37 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for one min. Cycling parameters for DDC were a 5-min soak at 94°C and 37 cycles of 94°C for 30 s, 58–63°C for 30 s and 72°C for one min. Cycling parameters for PEPCK were a 5-min soak at 94°C and 37 cycles of 94°C for 30 s, 55–57°C for 30 s and 72°C for one min. Cycling parameters for 12S were a 5-min soak at 94°C and 37 cycles of 94°C for 30 s, 62°C for 30 s and 72°C for one min. Cycling parameters for ND2 were a 5-min soak at 94°C and 37 cycles of 94°C for 30 s, 62°C for 30 s and 72°C for one min. The primers used to amplify ND2 are located in tRNA-Met and tRNA-Trp, respectively, which immediately flank this gene in vertebrates (Anderson *et al.*, 1982) and all insects studied thus far (Clary & Wolstenholme, 1985; Beard *et al.*, 1993; Mitchell *et al.*, 1993; Flook *et al.*, 1995), except *Apis mellifera* L. (Crozier & Crozier, 1993).

Table 1. Name, voucher number, life stage and source of species examined in this study and GenBank accession numbers for sequences obtained from them. Outgroups are arranged in order of phylogenetic relationships *sensu* Wood & Borkent (1989). Simuliid genera are arranged in alphabetical order below level of tribe as defined by Currie (1988). Genera in quotations are inappropriately assigned Australian species. L=larva, P=pupa, F=female, M=male.

Ref. no.	Taxon	Voucher no.	Life stage	Origin/source	28S D1-3	28S D9-10	EF-1 α	PEPCK	DDC	12S	ND2
Drosophilidae											
1	<i>Drosophila melanogaster</i> Meigen	–	A	GenBank database				Y00402	X04661		
Ptychopteridae											
2	<i>Bittacomorpha clavipes</i> (Fabricius)	JKM-1	A	USA: South Carolina						AF049468	
Dixidae											
3	<i>Dixa</i> sp.	JKM-2	L	USA: Arizona						AF049469	
Culicidae											
4	<i>Aedes aegypti</i> (Linnaeus)	–	F	GenBank database					U27581		
5	<i>Aedes albopictus</i> (Skuse)	–	A	GenBank database	L22060	L22060					
6	<i>Anopheles gambiae</i> Giles	–	L	GenBank database						L20934	L20934
Thaumaleidae											
7	<i>Androprosopa gillespieae</i> (Arnaud & Boussy)	JKM-3	L	USA: Arizona	AF007297	AF007298	AF003552	AF047531	AF078875	AF049471	AF078953
Chironomidae											
8	<i>Cardiocladius</i> sp.	JKM-4	L	USA: Arizona	AF007295	AF007296	AF003553	AF047530	AF078876	AF049470	
Simuliidae											
Parasimuliinae*											
9	<i>Parasimulium crosskeyi</i> Peterson	JKM-5	M	USA: Oregon	AF007299	AF007300	AF003554	AF047532	AF078877	AF049472	AF078954
Simuliinae*											
Prosimuliini**											
10	<i>Gymnopais fimbriatus</i> Wood	JKM-6	L	USA: Alaska	AF007309	AF007310	AF003559	AF047535	AF078878	AF049473	
11	<i>Helodon onychodactylus</i> Dyar & Shannon	JKM-7	L	USA: Arizona	AF007305	AF007306	AF003557	AF047534	AF078879	AF049474	AF078955
12	<i>Prosimulium formosum</i> Shewell	JKM-8	L	USA: Arizona	AF007301	AF007302	AF003555	AF047533	AF078880	AF049475	
13	<i>Prosimulium impostor</i> Peterson	JKM-9	L	USA: Arizona					AF078881		
14	<i>Twinnia nova</i> Dyar & Shannon	JKM-10	L	USA: Utah	AF007307	AF007308	AF003558		AF078882	AF049476	
15	<i>Urosimulium aculeatum</i> Rivosecchi	JKM-11	L	Spain	AF007303	AF007304	AF003556		AF078883		AF078956
Simuliini**											
16	<i>Austrosimulium bancrofti</i> Taylor cpx	JKM-12	L	Australia: Victoria	AF007311	AF007312	AF003560	AF047536	AF078884		AF078957
17	<i>Austrosimulium bancrofti</i> Taylor (Ipswich A1)	–	?	GenBank database						L02386	
18	<i>Austrosimulium mirabile</i> Mackerras & Mackerras	JKM-13	L	Australia: Queensland	AF007313	AF007314	AF003561				
19	' <i>Austrosimulium</i> ' <i>colboi</i> Davies & Gyorkos	JKM-14	L	Australia: Victoria	AF007349	AF007350					
20	<i>Cnephia ornithophilina</i> Davies, Peterson, & Wood	JKM-15	L	USA: South Carolina	AF007359	AF007360	AF003577			AF049477	AF078958
21	' <i>Cnephia</i> ' <i>aurantiacum</i> Tonnoir	JKM-16	L	Australia: Victoria	AF007317	AF007318	AF003563				
22	' <i>Cnephia</i> ' <i>strenua</i> Mackerras & Mackerras	JKM-17	P	Australia: Queensland	AF007319	AF007320	AF003564	AF047537			
23	' <i>Cnephia</i> ' <i>tonnoiri</i> Drummond	JKM-18	L	Australia: WA	AF007321	AF007322	AF003565				
24	' <i>Cnephia</i> ' <i>umbratorum</i> Tonnoir	JKM-19	F	Australia: Victoria	AF007343	AF007344	AF003575				
25	' <i>Cnephia</i> ' 'x'	JKM-20	P	Australia: WA	AF007333	AF007334	AF003588				
26	' <i>Cnephia</i> ' 'y'	JKM-21	P	Australia: WA	AF007335	AF007336	AF003586				
27	' <i>Cnephia</i> ' sp. nr <i>terebrans</i> Tonnoir (= 'GKW2')	JKM-22	L	Australia: WA	AF007337	AF007338	AF003587				
28	' <i>Cnephia</i> ' 'S. x. (east)'	JKM-23	L	Australia: Victoria	AF007339	AF007340	AF003584				
29	' <i>Cnephia</i> ' 'S. x. (west)'	JKM-24	L	Australia: WA	AF007341	AF007342	AF003585				
30	' <i>Cnephia</i> ' <i>pilfreyi</i> Davies & Gyorkos	JKM-25	L	Australia: WA	AF007345	AF007346	AF003583				

Table 1. continued.

Ref. no.	Taxon	Voucher no.	Life stage	Origin/source	28S D1-3	28S D9-10	EF-1 α	PEPCK	DDC	12S	ND2
31	<i>Cnesia dissimilis</i> Edwards	JKM-26	L	Argentina	AF007331	AF007332	AF003568	AF047538		AF049478	AF078959
32	<i>Crozetia crozetensis</i> Womersley	JKM-27	L	Isle de Crozet	AF007315	AF007316	AF003562		AF078885		
33	<i>Ectemnia reclusa</i> Moulton & Adler	JKM-28	L	USA: South Carolina	AF007361	AF007362	AF003589			AF049479	
34	<i>Greniera denaria</i> Davies, Peterson & Wood	JKM-29	L	Canada: Ontario	AF007355	AF007356	AF003573			AF049480	
35	<i>Greniera fabri</i> Doby & David	JKM-30	L	Spain	AF007357	AF007358	AF003574	AF047539	AF078886		
36	<i>Gigantodax adleri</i> Moulton	JKM-31	L	USA: Arizona	AF007327	AF007328	AF003566				
37	<i>Gigantodax marginalis</i> Edwards	JKM-32	L	Argentina	AF007329	AF007330	AF003567				
38	<i>Mayacnephia</i> sp. nr <i>osborni</i> Stains & Knowlton	JKM-33	L	USA: Arizona	AF007353	AF007354	AF003572			AF049481	
39	<i>Metacnephia sommermanae</i> Stone	JKM-34	L	USA: Alaska	AF007363	AF007364	AF003578	AF047540		AF049482	
40	<i>Paracnephia thornei</i> de Meillon	JKM-35	L	Rep. of South Africa	AF007323	AF007324	AF003569		AF078887		
41	<i>Paraustrosimulium anthracinum</i> Moore	JKM-36	L	Chile	AF007347	AF007348	AF003576			AF049483	
42	<i>Procnephia rhodesiense</i> (Crosskey)	JKM-37	L	Zimbabwe	AF007325	AF007326	AF003570				
43	<i>Simulium (Edwardsellum) sirbanum</i> Vajime & Dunbar	JKM-38	L	Liberia	AF007371	AF007372	AF003581				
44	<i>Simulium (Hellichiella) canonicolum</i> (Dyar & Shannon)	JKM-39	L	USA: Arizona	AF007367	AF007368					
45	<i>Simulium (Hellichiella) congareenarum</i> Dyar & Shannon	JKM-40	L	USA: South Carolina	AF007365	AF007366	AF003579		AF078888		AF078960
46	<i>Simulium (Hellichiella) curriei</i> Adler & Wood	JKM-41	L	USA: Arizona						AF049484	
47	<i>Simulium (Nevermannia) pugetense</i> Dyar & Shannon cpx	JKM-42	L	USA: Arizona	AF007369	AF007370	AF003580				
48	<i>Simulium (Psilozia) encisoii</i> Vargas & Díaz Nájera	JKM-43	P	USA: Arizona				AF047531			
49	<i>Simulium (Simulium) reptans</i> Linnaeus	JKM-44	F	UK: England	AF007373	AF007374	AF003582				
50	<i>Stegopterna</i> 'W' (Madahar, 1969)	JKM-45	L	USA: West Virginia	AF007351	AF007352	AF003571			AF049485	

Sensu* Crosskey (1988) and Currie (1988).*Sensu* Currie (1988).

DNA clean-up and sequencing

PCR products intended for direct sequencing were cleaned and concentrated using Microcon-100 microconcentrators (Amicon Incorporated, Beverly, Massachusetts) and resuspended in deionized water. PCR fragments of EF-1 α , DDC and PEPCK for some taxa were isolated and amplified via T/A cloning (Invitrogen Corporation, Carlsbad, California). Plasmid DNA was purified using Wizard Minipreps (Promega Corporation, Madison, Wisconsin).

DNA sequencing was performed at the University of Arizona's Macromolecular Sequencing Facility (Arizona Research Laboratories), using an upgraded ABI 373 or ABI 377 automated sequencer. Both strands of all DNA fragments, except those of ND2, were sequenced. The entire EF-1 α fragment was sequenced using the sequencing primers listed in Table 3 and the M13 priming sites of the PCR primers, if directly sequenced, or the sequencing primers and the SP6 and T7 priming sites of the plasmid vector, if cloned. Regions encompassing D1-D3 and D9-D10 of 28S were sequenced using D1 SP and D3 SP and D9 SP and D10 SP, respectively. DDC and PEPCK, if cloned, were sequenced using the M13 priming sites of the plasmid or, if directly sequenced, using DDC 5' SP and DDC 3' SP or 150dF and 1623drc, respectively. The single strand of ND2 sequenced was obtained using ND2 5' PCR. Chromatograms from sequencing reactions were assembled and ambiguous and conflicting bases resolved using Sequencher 2.1 (Gene Codes Corporation, Ann Arbor, Michigan).

DNA alignment

Ideally, alignment of 12S and 28S would be inferred from secondary structure. However, preliminary secondary structure inferences using Mulfold 2.0d82 (Zuker & Jaeger, 1990) and RNAFOLD (University of Wisconsin Genetics Computer Group software package) yielded outputs that were difficult to reconcile. Inference of secondary structure was not more ambitiously pursued because: (1) 12S sequences differed only by numerous small insertion/deletion events, (2) deletion of regions in 28S that were difficult to align manually had little effect upon either cladogram topology or node support, (3) although rates of compensatory change in stem regions may be high, they often are delayed more than 45 million years (Kraus *et al.*, 1992), which means that in a rapidly evolving group many will potentially be informative at higher taxonomic levels and therefore should not be downweighted, (4) there is no agreement regarding how much known compensatory changes and sites in loop structures should be weighted (Wheeler & Honeycutt, 1988; Dixon & Hillis, 1991; Bruns *et al.*, 1992; Bakker *et al.*, 1994; Miyamoto *et al.*, 1994) and (5) applying such schemes often has little effect upon resolution (Smith, 1989; Bruns *et al.*, 1992; Hedges *et al.*, 1996; Hickson *et al.*, 1996).

Portions of the 28S and 12S data sets were difficult to align manually without potentially introducing bias. As a result,

initial alignments of 28S and 12S sequences were done using Clustal W (Thompson *et al.*, 1996); the default settings of this program, i.e. transitions weighted, gap creation/extension costs of 10:5, were used. Slight optimization of these alignments was done manually using SeqApp 1.9a157 (Gilbert, 1992).

Two methods were used in order to determine the effects of alignment on phylogenetic inferences from 28S and 12S sequences. For 28S, two alignments were created. One alignment contains sequences from a culicid, chironomid, thaumaleid and all simuliids and is the one used to infer relationships outside of Simuliidae. This matrix (28S matrix 1) was additionally analysed with several difficult-to-align regions, e.g. nucleotides 423–460, 473–480, 526–560, 628–746, 832–850, 870–885, 950–996, 1244–1281 and 1462–1514, excluded. The other (28S matrix 2) excludes the mosquito and chironomid, which were considerably more difficult to align with simuliids than was the thaumaleid. The latter matrix is a less ambiguous alignment from which to infer basal relationships within Simuliidae. For 12S, the elision method (Wheeler *et al.*, 1995) was used. Owing to the small size of the fragment sequenced, this method seemed favourable to excluding variation. The following gap creation/extension costs were used: 2/1, 5/1, 5/2; 5/10, 7/2; 10/2, 10/5; 20/5, 20/10 and 40/1. These ten alignments were combined and analysed simultaneously in addition to analyses of the manual alignment.

Of the three nuclear protein coding genes examined, only PEPCK was without introns in all of the taxa examined. EF-1 α was without introns for all taxa except *Ectemnia reclusa* Moulton & Adler, which has an intron of seventy bases in length positioned between sites 799 and 800. Because this intron position is not homologous to any in the F2 copy of EF-1 α in *Drosophila* and direct sequencing of the PCR product yielded no ambiguities in this region, this copy is presumed orthologous with the others.

The DDC sequences contained either one or two introns. The thaumaleid and all simuliids other than *Simulium congareenarum* have a pair of introns located in identical positions. The proximal intron is absent in *Cardiocladius* sp. and the distal intron is absent in *Simulium congareenarum*. The second intron in *Cardiocladius* sp. is unique among these taxa and begins only thirteen bases downstream of the priming site of DDC 5' PCR. The proximal intron ranged in size from fifty-eight bases in *Austrosimulium bancrofti* to 557 bases in *Parasimulium crosskeyi* Peterson. The distal intron ranged from fifty-six bases in *Austrosimulium bancrofti* to seventy-one bases in *Prosimulium formosum* Shewell and *Urosimulium aculeatum* Rivosecchi. Both DDC introns were difficult to impossible to align among genera, whereas between the two species of *Prosimulium*, each of which represents a major species group (Rothfels, 1979, 1981; Crosskey, 1988), both were easily aligned and exhibit considerable conservation.

Introns were deleted to facilitate alignment and translation into protein, the latter of which was done using MacClade 3.04 (Maddison & Maddison, 1992). After intron removal, alignment of EF-1 α and PEPCK sequences was straightforward, i.e. no gaps were needed. In DDC alignments, gaps were needed to

compensate for the absence of a single codon in *Drosophila melanogaster* with respect to the other taxa examined. Introns were not considered in phylogenetic analyses.

All data matrices analysed in this work, except the 12S elision matrix, appear in Moulton (1997). These can be obtained from the author, upon request, as NEXUS- or NBRF-formatted files.

Sequence analysis

Phylogenetic analysis was done using PAUP*d54 and d56 (Swofford, 1996–1997; used with author's permission). Uncorrected pairwise distances, calculations of and tests for differences in nucleotide composition and all phylogenetic inferences, except the likelihood analysis of 28S sequences as aligned in data matrix, were calculated using PAUP*d52 (Swofford, 1996; used with author's permission). The likelihood analysis of 28S sequences was done using PAUP*d54. Ranges of pairwise distances are presented for molecular sequence data (sometimes partitioned into discrete character types) in order to give an impression of their variabilities and phylogenetic utilities within and among previously recognized taxonomic groups. Nucleotide frequencies were calculated for these sequences or partitions of them to discover potential shifts in base composition, which are known to be detrimental to the accuracy of phylogenetic inference (Hasegawa & Hashimoto, 1993; Steel *et al.*, 1993; Collins *et al.*, 1994).

For each combination of data matrix, constraints and weights (parsimony only), most parsimonious, minimum evolution and neighbour-joining cladograms were sought. Maximum likelihood trees were sought only from the manual alignment of 28S sequences. All searches, except parsimony and minimum evolution bootstrap analyses of the elision matrix of 12S, used PAUP*'s heuristic search option, with TBR branch rearrangement; nearest neighbour interchanges were performed in the one exception. Characters were treated as unordered in all parsimony analyses and MAXTREES was set to increase incrementally. Parsimonious cladograms were found by conducting 200 searches, with each search begun from cladograms acquired by stepwise addition with a random addition sequence order. Each island of most parsimonious cladograms was found at least fifteen times using this approach. Minimum evolution cladograms were found by conducting a single search from a cladogram acquired by neighbour-joining. HKY85 (Hasegawa *et al.*, 1985) corrected distances were used in distance analyses of nucleotide sequences, and mean distances were used in those of amino acid sequences.

The tree of highest likelihood from the manually aligned matrix of 28S sequences was found by using a single search beginning with the minimum evolution cladogram from these same data. The HKY85 substitution model of nucleotide evolution was assumed, with rate heterogeneity accommodated using a discrete approximation with four rate categories to a gamma distribution with the shape parameter equal to five.

Robustness of nodes was evaluated by bootstrap resampling and through the calculation of decay indices (Bremer, 1988). Bootstrap values for parsimony, minimum evolution and neighbour-joining analyses were calculated with 200 bootstrap replicates; for parsimony bootstrapping, each replicate consisted of a single search starting with a cladogram built by stepwise addition using the simple addition sequence, whereas for distance bootstrapping, each replicate was begun from the neighbour-joining cladogram. A MAXTREES of 200 was imposed on all bootstrap analyses. Decay indices were calculated for particular clades by constraining PAUP*'s search to find the most parsimonious cladograms without the group, which was accomplished by forcing the particular group not to be monophyletic using a constraint cladogram, and then comparing the cladogram lengths of these cladograms to the most parsimonious cladograms. Most parsimonious cladograms in which particular clades are not present were found by conducting twenty-five searches using TBR with each search begun from cladograms acquired by stepwise addition with a random addition sequence order.

Differential weighting of codon positions considering both type of amino acid replacement (e.g. conservative vs. non-conservative) and domain structure (e.g. transmembrane vs. DNA-binding) probably represents the best means to extract phylogenetic information from protein coding genes. Because domain structure is largely unknown for the genes examined here, differential weighting considering only codon position was conducted. The codon weighting schemes used in analyses of DDC, PEPCK, and ND2 nucleotides were 1/1/1, 1/1/0, 3/5/1, 3/10/1 and 3/20/1. The choice of weights is entirely exploratory; however, the downweighting of third positions relative to the others and the higher weight given to second positions relative to first positions follows logic, given our knowledge of the evolution of codon positions.

The following regions of questionable alignment in 28S data matrix one were excluded from some phylogenetic analyses: 423–460, 473–480, 526–560, 628–746, 832–850, 870–885, 950–996, 1244–1281 and 1462–1514. Although secondary structure inference was not carried out, deleting the aforementioned regions is intended to serve as a proxy for the exclusion of loops and highly variable stem/loop structures. The results are inferred to provide a more cautious hypothesis of relationships.

Due to limited taxon sampling outside of Simuliidae and consideration of only its basalmost bifurcations, I have taken the following approach in presenting results from the numerous phylogenetic analyses conducted on these data: (1) only basal clades exhibiting bootstrap scores of 50% or more in at least some analyses of each locus are considered and (2) results are shown in table form to provide a straightforward means to observe the effects taxon sampling and differential weighting had on the recovery of certain clades. The former action was taken to provide conservative estimates of relationships. All 50% majority rule consensus cladograms summarizing results of bootstrap analyses are concordant with the most parsimonious, minimum distance

Table 2. Primers used in the study.

Name	Direction	Position	Sequence (5'→3')								
28S		5									
5.8S PCR	Fwd	2796	CTA	ACT	GCG	TGT	TGT	CAT	GTG	AAC	TGC
D1 SP	Fwd	3363	GGG	AGG	AAA	AGA	AAC	TAA	C		
D3 PCR	Fwd	4077	TTG	AAA	CAC	GGA	CCA	AGG	AGT	CTA	
D3 SP	Rev	4267	GCA	TAG	TTC	ACC	ATC	TTT	C		
D6 SP	Fwd	4749	CGC	TAA	GGA	GTG	TGT	AAC			
D8 SP	Rev	5826	GCA	CTG	GGC	AGA	AAT	CA			
D9 SP	Fwd	5933	AGC	CAA	ATG	CCT	CGT	ATC			
D8 PCR	Rev	6039	TTA	GAG	TCA	AGC	TCA	AAA	GGG	TCT	TCT
D10 SP	Rev	6404	TAC	CGC	CCC	AGT	CAA	AC			
D10 PCR	Rev	6406	TGA	GAG	ATG	TAC	CGC	CCC	AGT	CAA	
DDC		6									
DDC 5' PCR	Fwd	334	ATG	CCN	GGN	GTN	CAN	CAY	TGG		
DDC 5' SP	Fwd	405	ATT	GTK	GCY	GAY	ATG	TTG	AG		
DDC 3' SP	Rev	983	SGW	TTC	RAT	GCC	YTT	CAT			
DDC 3' PCR	Rev	1013	CAT	CCA	YTT	RTG	NGG	RTT	RAA	RTT	
PEPCK		7									
284dF ¹	Fwd	901	ATN	GTN	CAY	YTN	TTY	GAR	TGG		
18.5dF ¹	Fwd	991	TGT	GGN	AAR	ACC	AAV	YTG	GCC	ATG	
150dF	Fwd	1052	GTY	GGW	GAY	GAY	ATY	GCN	TGG	ATG	A
1623drc	Rev	1623	ART	TGA	CRT	GGA	AKA	TYT	TTG	GTG	
22.5drc ¹	Rev	1630	GAA	CCA	RTT	RAC	RTG	RAA	GAT	C	
1691drc	Rev	1706	CCN	GAN	ATN	ACR	TCR	CAR	TA		
EF-1 α		8									
M3 PCR ^{2,3}	Fwd	2103	CAG	GAA	ACA	GCT	ATG	ACC	CAC	ATY	AAC
			ATT	GTC	GTS	ATY	GG				
EF-1 α 5' SP	Fwd	2085	TGT	TTA	CAA	AAT	TGG	CGG	TAT		
Sam	Fwd	2429	YGA	TTG	TCG	CCG	CCG	GTA	CTG	GTG	AAT
Joe-2	Fwd	2704	CCG	TGG	TWC	AAG	GGA	TGG			
Joe	Fwd	2706	CHT	GGT	WCA	AGG	GAT	GGA	A		
Shemp	Rev	2822	TCC	RAT	ACC	NCC	RAT	TTT	GTA		
Curly	Rev	2831	GTA	CTG	TTC	CGA	TAC	CGC	C		
EF-1 α 3' SP	Rev	3363	TTC	CAA	TAC	CGC	CAA	TTT	TG		
rcM4 PCR ^{2,3}	Rev	3344	TGT	AAA	ACG	ACG	GCC	AGT	ACA	GCV	ACK
			GTY	TGY	CTC	ATR	TC				
12S		9									
12Sbi ⁴	Fwd	14233	AAG	AGC	GAC	GGG	CGA	TGT	GT		
12Sair ⁴	Fwd	14612	AGG	GTA	TCT	AAT	CCT	AGT	TT		
12Sai ⁴	Rev	14588	AAA	CTA	GGA	TTA	GAT	ACC	CTA	TTA	T
12Sz	Rev	14759	AGT	ATT	GGT	AAA	ATT	TGT	GCC	AGC	
ND2		9									
ND2 5'	Fwd	209	AAG	CTA	YTG	GGT	TCA	TAC	C		
ND2 3'	Rev	1442	ART	GGC	TGA	AGT	TTA	GGC	GAT	A	

¹Designed by Friedlander *et al.* (1996).²Designed by Cho *et al.* (1995).³The initial eighteen bases of these primers correspond to M13R and M13F primer sequences, respectively.⁴Designed by T. Kocher (Kocher *et al.*, 1989) and/or C. Simon (Simon *et al.*, 1994).⁵Position of 3' base in rRNA gene cluster of *Drosophila melanogaster* (GenBank accession M21017).⁶Position of 3' base in *Aedes aegypti* dopa decarboxylase (DDC) gene sequence (GenBank accession U27581).⁷Position of 3' base in *Drosophila melanogaster* phosphoenolpyruvate carboxykinase (PEPCK) gene sequence (GenBank accession Y00402).⁸Position of 3' base in the coding region of the *Drosophila melanogaster* F1 EF-1 α gene sequence (Walldorf *et al.*, 1985) (GenBank accession M11744).⁹Position of 3' base in *Drosophila yakuba* complete mtDNA sequence (GenBank accession X03240).

and neighbour-joining cladograms found using identical parameters. Higher-level relationships within Simuliidae will

be treated elsewhere (J. K. Moulton, unpublished observations).

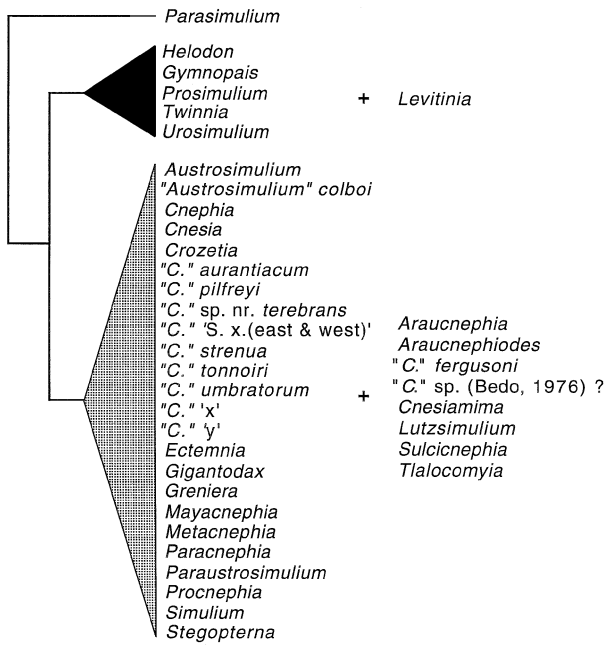


Fig. 1. Cladogram depicting basal lineages within Simuliidae as inferred from molecular (left) and concordant morphological (right) evidence, the latter presented by Currie (1988). Bootstrap scores, decay values and well supported alternative topologies, if present, corresponding to or inferred from molecular data presented herein appear in Tables 4–7. Intratribal relationships inferred from molecular data with significant bootstrap support not shown.

Results and discussion

Properties of the data sets

28S sequences are presented from forty species, thirty-eight of which are simuliids (Table 2). These sequences range in length from 1384 bases in *Cardiocladius* sp. to 1506 bases in *Aedes albopictus* (Skuse). Data matrix 1 contains 1613 aligned characters, of which 700 (43%) are invariant, 483 (30%) are parsimony uninformative and 430 (27%) are parsimony informative. Data matrix 2 contains 1556 aligned characters, of which 1050 (67%) are invariant, 273 (18%) are parsimony uninformative and 233 (15%) are parsimony informative. Base composition between *Cardiocladius* sp., *Androsopoda gillespieae* and the simuliids does not vary significantly ($\chi^2 = 54.5$, d.f. = 117, $P > 0.9999$), but it does when *Aedes albopictus* is included ($\chi^2 = 276$, d.f. = 120, $P < 0.0001$). The average frequency of each nucleotide among all of these taxa, excluding *Aedes albopictus*, is as follows: A (31%), C (16%), G (22%) and T (31%). The same values for *Aedes albopictus* are 23%, 27%, 29% and 21%, respectively.

EF-1 α sequences were obtained from thirty-eight species, thirty-six of which are simuliids (Table 2). The nucleotide data matrix comprises 1239 aligned sites, of which 751 (60.6%) are invariant, 82 (7%) are parsimony uninformative and 406 (32.8%) are parsimony informative. The amino acid data matrix comprises 413 aligned sites, of which 334 (80.9%) are

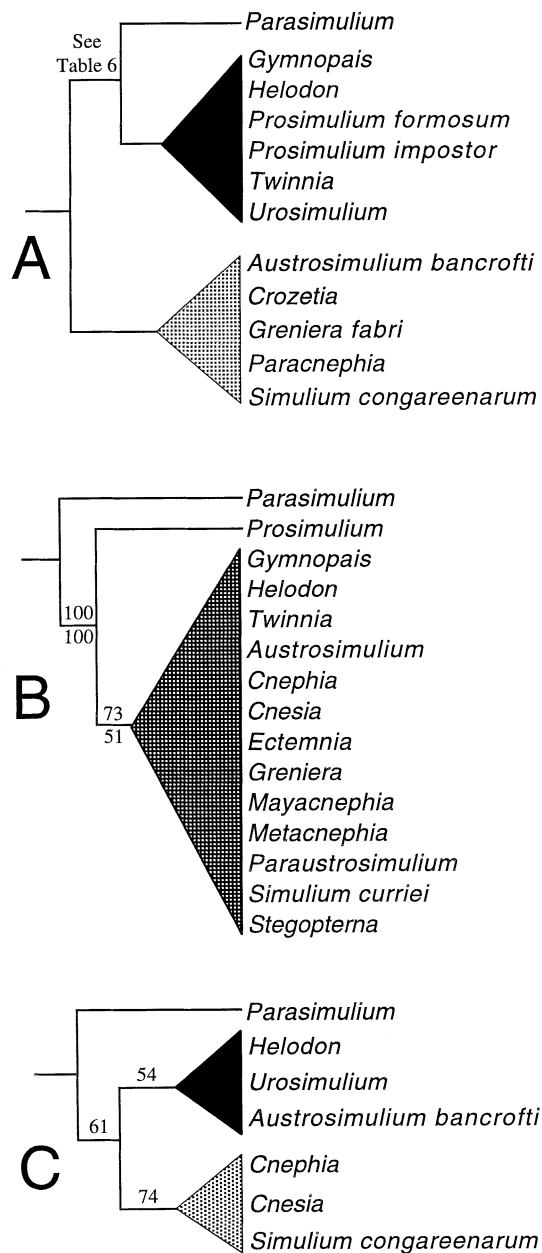


Fig. 2. Topologies depicting basal relationships within Simuliidae differing from that of Currie (1988) and most analyses of molecular data presented herein. A, 1/1/1 and 3/5/1 codon weighting parsimony analyses of DDC nucleotides, minimum evolution (ME) and neighbour-joining (NBJ) analyses of all nucleotides, and NBJ analysis of DDC amino acids. B, Parsimony and NBJ analyses of 12S elision matrix with *Dixa* as distal outgroup, i.e. *Bittacomorpha* excluded. C, Parsimony analysis of ND2 amino acids.

invariant, 44 (10.6%) are parsimony uninformative and 35 (8.5%) are parsimony informative. When nucleotide composition is analysed by codon position, first ($\chi^2 = 9.32$, d.f. = 111, $P > 0.9995$) and second ($\chi^2 = 2.17$, d.f. = 111, $P > 0.9995$) positions do not vary significantly, but third positions exhibit notable differences ($\chi^2 = 584$, d.f. = 111, $P < 0.0001$). Third

Table 3. Statistics of results from phylogenetic analyses of 28S, EF-1 α , DDC, PEPCK, 12S and ND2 data sets.

Locus/analysis	No. MPTs	No. islands	Parsimony TL	CI	RI	RC	HI	Distance ME score
28S								
data matrix 1	5	1	1928	0.70	0.60	0.42	0.30	1.37764
data matrix 1 ^a	24	1	1002	0.75	0.66	0.50	0.25	0.94872
data matrix 2	36	1	1044	0.60	0.62	0.37	0.40	0.69886
Elision matrix	1	1	10862	0.59	0.60	0.36	0.41	0.71951
EF-1 α								
all taxa included	1	1	2543	0.30	0.45	0.14	0.70	1.88035
<i>Cardiocladius</i> excluded	1	1	2398	0.31	0.46	0.14	0.69	1.88197
PEPCK ^b								
DNA – 1/1/1	3	2	820	0.54	0.40	0.22	0.46	1.47025
DNA – 1/1/0	6	1	183	0.69	0.53	0.36	0.31	0.47751
DNA – 1/2/1	5	2	882	0.55	0.41	0.23	0.45	NA
DNA – 3/5/1	1	1	1311	0.60	0.45	0.27	0.40	NA
DNA – 3/10/1	1	1	1604	0.62	0.48	0.30	0.38	NA
DNA – 3/20/1	1	1	2184	0.65	0.52	0.34	0.35	NA
AA	1	1	138	0.76	0.55	0.42	0.24	0.69064
PEPCK ^c								
DNA – 1/1/1	3	2	720	0.56	0.41	0.23	0.44	1.27939
DNA – 1/1/0	12	1	151	0.70	0.53	0.37	0.30	0.39229
DNA – 1/2/1	1	1	764	0.57	0.43	0.24	0.43	NA
DNA – 3/5/1	2	2	1112	0.61	0.46	0.28	0.39	NA
DNA – 3/10/1	1	1	1330	0.64	0.50	0.32	0.37	NA
DNA – 3/20/1	1	1	1760	0.67	0.55	0.37	0.33	NA
AA	1	1	76	0.76	0.57	0.43	0.24	0.57426
DDC ^b								
DNA – 1/1/1	3	2	975	0.50	0.44	0.22	0.50	1.85026
DNA – 1/1/0	12	1	226	0.65	0.60	0.39	0.35	0.62639
DNA – 1/2/1	1	1	1043	0.51	0.45	0.23	0.49	NA
DNA – 3/5/1	1	1	1573	0.56	0.49	0.28	0.44	NA
DNA – 3/10/1	1	1	1898	0.60	0.52	0.31	0.40	NA
DNA – 3/20/1	1	1	2548	0.65	0.56	0.36	0.35	NA
AA	144	1	160	0.81	0.68	0.56	0.19	0.78515
DDC ^c								
DNA – 1/1/1	1	1	869	0.52	0.45	0.23	0.48	1.63747
DNA – 1/1/0	18	1	186	0.66	0.61	0.40	0.34	0.51707
DNA – 1/2/1	1	1	923	0.54	0.46	0.25	0.46	NA
DNA – 3/5/1	3	2	1355	0.58	0.50	0.30	0.42	NA
DNA – 3/10/1	1	1	1619	0.63	0.54	0.34	0.37	NA
DNA – 3/20/1	1	1	2139	0.69	0.59	0.41	0.32	NA
AA	373	1	130	0.82	0.73	0.60	0.18	0.65044
12S								
all taxa included	1	1	708	0.54	0.40	0.21	0.46	1.34670
<i>Bittacomorpha</i> excluded	2	1	643	0.56	0.39	0.22	0.44	1.22923
<i>Bittacomorpha</i> & <i>Dixa</i> excluded	1	1	603	0.57	0.37	0.21	0.43	1.15017
<i>Bittacomorpha</i> , <i>Dixa</i> & <i>Anopheles</i> excluded	4	3	540	0.59	0.37	0.22	0.41	1.03323
12S Elision								
all taxa included	2	1	7098	0.55	0.38	0.21	0.45	1.34204
<i>Bittacomorpha</i> excluded	2	1	6377	0.57	0.38	0.22	0.43	1.21011
<i>Bittacomorpha</i> & <i>Dixa</i> excluded	2	1	5959	0.59	0.36	0.21	0.41	1.12865
<i>Bittacomorpha</i> , <i>Dixa</i> & <i>Anopheles</i> excluded	1	1	5357	0.61	0.36	0.21	0.39	1.02125
ND2								
DNA – 1/1/1	1	1	520	0.67	0.31	0.20	0.34	1.24315
DNA – 1/1/0	3	1	224	0.74	0.42	0.31	0.26	0.75952
DNA – 3/5/1	1	1	1126	0.71	0.36	0.26	0.29	NA
DNA – 3/10/1	1	1	1513	0.74	0.39	0.28	0.27	NA
DNA – 3/20/1	1	1	2273	0.76	0.43	0.33	0.24	NA
AA	1	1	200	0.85	0.52	0.45	0.15	1.10769

^aHypervariable regions excluded (= sites 423–460, 473–480, 526–560, 628–746, 832–850, 870–885, 950–996, 1244–1281 and 1462–1514).^bAll taxa included.^c*Drosophila* excluded.

position base composition in the chironomid is more similar to that of Simuliinae than is *Parasimulium crosskeyi* and significant differences exist even within *Simulium* ($\chi^2=29.6$, d.f.=9, $P<0.0005$).

PEPCK sequences are presented for eleven species, nine of which are simuliids representing as many genera (Table 2). The nucleotide data matrix comprises 543 aligned sites, of which 297 (55%) are invariant, fifty-eight (11%) are parsimony uninformative and 188 (34%) are parsimony informative. The amino acid data matrix comprises 181 aligned characters, of which 123 (68%) are invariant, twenty-five (14%) are parsimony uninformative and thirty-three (18%) are parsimony informative. Nucleotide composition at first ($\chi^2=4.12$, d.f.=36, $P>0.9995$) and second ($\chi^2=3.58$, d.f.=36, $P>0.9995$) positions do not vary significantly among these taxa. At the third position, however, nucleotide composition varies considerably ($\chi^2=182$, d.f.=36, $P<0.0001$), even among simuliids ($\chi^2=130$, d.f.=27, $P<0.0001$).

DDC sequences are presented for fourteen species representing eleven genera of black flies and two outgroup taxa (Table 2). The nucleotide data matrix comprises 519 aligned characters, of which 239 (46%) are invariant, fifty-six (11%) are parsimony uninformative and 224 (43%) are parsimony informative. The amino acid data matrix comprises 173 aligned characters, of which 102 (59%) are invariant, forty (23%) characters are parsimony uninformative and thirty-one (18%) are parsimony informative. Nucleotide composition at first ($\chi^2=34.12$, d.f.=45, $P>0.8816$) and second ($\chi^2=2.38$, d.f.=45, $P>0.9995$) positions do not vary significantly among these taxa. Third positions, however, vary considerably ($\chi^2=245$, d.f.=45, $P<0.0001$), even among simuliids ($\chi^2=166$, d.f.=33, $P<0.0001$).

12S rDNA sequences are presented from seventeen species, fourteen species of simuliids representing as many genera and three outgroup taxa (Table 2). Additional sequences, i.e. *Anopheles gambiae* and *Austrosimulium bancrofti*, were obtained from GenBank. The data matrix comprises 531 aligned nucleotides, of which 262 (49%) are invariant, 102 (19%) are parsimony uninformative and 167 (32%) are parsimony informative. Nucleotide composition across all taxa does not vary significantly ($\chi^2=48.0$, d.f.=60, $P>0.8680$), but as expected from previous studies, e.g. Ballard *et al.* (1992) and Ballard (1994), these sequences are heavily skewed towards adenine and thymine (mean=78%). Mean frequencies of adenine, thymine, cytosine and guanine, respectively, across all taxa are 40, 38, 14 and 8%.

ND2 sequences are presented for eight species, seven of which represent simuliid genera and the other a thaumaleid (Table 2). The nucleotide data matrix comprises 454 aligned characters, of which 207 (46%) are invariant, 106 (23%) are parsimony uninformative and 141 (31%) are parsimony informative. The amino acid data matrix comprises 152 aligned characters, of which sixty-three (42%) are invariant, forty-eight (32%) are parsimony uninformative and forty (26%) are parsimony informative. As with 12S, these sequences are heavily skewed towards adenine and thymine.

Across all taxa, mean frequencies of adenine, thymine, cytosine and guanine, respectively, were 33%, 44%, 13% and 10%. Nucleotide composition at first ($\chi^2=12.2$, d.f.=24, $P>0.9778$) and second ($\chi^2=6.05$, d.f.=24, $P>0.9999$) positions do not vary significantly among these taxa. Third positions, however, vary considerably ($\chi^2=36.1$, d.f.=24, $P<0.0541$), although less so among simuliids alone ($\chi^2=21.4$, d.f.=18, $P<0.2582$).

Phylogenetic utility of third codon positions

For the ages of divergence treated in this chapter, it is important to assess the phylogenetic utility of synonymous variation in the protein-coding genes examined herein. It is notable that the greatest pairwise divergences at the third codon position in the EF-1 α (55% between *Parasimulium crosskeyi* and *Simulium congareenarum*), DDC (67% between *Parasimulium crosskeyi* and *Crozetia crozetensis*), PEPCK (63% between *Drosophila melanogaster* and *Parasimulium crosskeyi*) and ND2 (49% between *Androprosopa gillespieae* and *Urosimulium aculeatum*) data sets do not occur between the most distantly related pair of taxa examined, and except for that of EF-1 α , they do not even occur among the most distantly related simuliids (Moulton, 1997).

Drastic shifts in nucleotide composition can cause problems in phylogenetic inference (Hasegawa & Hashimoto, 1993; Steel *et al.*, 1993; Collins *et al.*, 1994). Analysis of nucleotide composition at third positions in the protein-coding genes examined reveal significant differences among these taxa, including numerous instances within Simuliidae. Because third positions exhibit divergences ranging from 20% to 40% among EF-1 α sequences and 30–50% or more among the other three protein coding genes, do not always increase with increasing divergence when mapped onto cladograms inferred as best estimates of relationships within Simuliidae (Moulton, 1997) and exhibit significant base compositional changes among taxa, they are deemed at or near saturation and less appropriate for inferring relationships above the level of genus than are first and second position non-synonymous changes. As a result, severe downweighting or exclusion of third positions appeared warranted.

Phylogenetic inferences

Figure 1 summarizes basal simuliid relationships as inferred from a majority of the phylogenetic analyses performed. Figure 2 shows three alternative scenarios as observed in a few analyses of 12S, DDC and ND2. Length, island number and character-fit indices of the most parsimonious and distances of the minimum length cladograms found are presented in Table 3. Tables 4, 5, 6 and 7 list, by analysis, bootstrap and decay values for the most commonly observed clades in cladograms inferred from 28S, EF-1 α ; PEPCK; DDC; and 12S, ND2; respectively.

Table 4. Summary of support for relationships between Chironomidae, Thaumaleidae and Simuliidae, and basal clades within Simuliidae as inferred from 28S and EF-1 α nucleotides. Numbers not in parentheses are bootstrap scores; numbers in parentheses are decay indices. 'NA' indicates no score was assigned to the node. '-' indicates the clade was present in fewer than 50% of the bootstrap replicates.

Locus/analysis	Clades of interest				
	Thaumaleidae + Simuliidae	Simuliidae	Simuliinae	Prosimuliini <i>sensu</i> Currie (1988)	Simuliini <i>sensu</i> Currie (1988)
28S (matrix 1)					
Parsimony ¹	100 (51)	100 (56)	99 (14)	99 (7)	100 (14)
Minimum evolution ¹	100	100	91	100	–
Neighbour-joining ¹	100	100	98	100	84
Parsimony ²	100 (44)	100 (30)	77 (9)	96 (3)	64 (9)
Minimum evolution ²	100	100	99	100	86
Neighbour-joining ²	100	100	99	100	94
28S (matrix 2)					
Parsimony	NA	NA	100 (21)	97 (8)	100 (13)
Minimum evolution	NA	NA	100	100	78
Neighbour-joining	NA	NA	100	100	90
EF-1 α					
Parsimony ³	NA	NA	54 (4)	99 (16)	71 (0)
Minimum evolution ³	NA	–	74	100	81
Neighbour-joining ³	NA	–	–	100	66
Parsimony ⁴	NA	NA	96 (12)	99 (17)	88 (10)
Minimum evolution ⁴	NA	NA	99	100	76
Neighbour-joining ⁴	NA	NA	96	100	72

¹All taxa included.

²Hypervariable regions excluded (= sites 423–460, 473–480, 526–560, 628–746, 832–850, 870–885, 950–996, 1244–1281 and 1462–1514).

³*Androprosopa* and *Cardiocladius* included.

⁴*Cardiocladius* excluded.

Cladogram evaluation with reference to previous hypotheses

Relationship between Chironomidae, Simuliidae and Thaumaleidae. Molecular studies to date have not been intense enough, with respect to both taxon and genomic sampling, to adequately address relationships within Culicomorpha. Using different portions of the same gene, 28S, Pawlowski *et al.* (1996) failed to recover Culicomorpha *sensu* Hennig (1973), whereas Friedrich & Tautz (1997) did so with strong statistical support. Neither Friedrich & Tautz (1997) nor Pawlowski *et al.* (1996), however, recovered Chironomoidea *sensu* Hennig (1973). Miller *et al.* (1997) likewise did not observe support for Chironomoidea from analyses of a fragment comprised of 5.8S and 18S. The study of Pawlowski *et al.* (1996) is unique because they included a thaumaleid. They observed strong decay (Bremer, 1988) and bootstrap support for a sister-group relationship between it and Simuliidae, a finding that is in disagreement with most views based upon morphological evidence (Hennig, 1973; Wood, 1978; Wood & Borkent, 1989; Oosterbroek & Courtney, 1995).

Although sparse taxon sampling outside of Simuliidae, including the absence of a ceratopogonid, prevents strong conclusions being drawn from these data, for the sake of completeness and to prompt additional, more intensive studies in this area, I include results from analyses in which resolution of relationships among these three taxa

was strongly supported. Resolution of this tritomy lends support to either the bulk of the morphological studies or the lone molecular study to date that included at least one chironomid or ceratopogonid, simuliid and thaumaleid, e.g. Pawlowski *et al.* (1996).

The Thaumaleidae + Simuliidae arrangement was significantly supported in all 28S (Table 4), most PEPCK (Table 5) and most 12S elision (Table 7) analyses. The Chironomidae + Simuliidae arrangement was significantly supported only in the 3/20/1 (nt1/nt2/nt3) codon weighting bootstrap analysis of DDC nucleotides with *Drosophila melanogaster* excluded. Perhaps additional investigations implementing more intensive taxon sampling and larger sequences, preferably from more appropriate loci, will be better able to resolve these issues.

Parasimulium as sister group of all other Simuliidae. This arrangement has been hypothesized numerous times from morphological evidence (Smart, 1945; Shewell, 1958; Grenier & Rageau, 1960; Dumbleton, 1963; Crosskey, 1969, 1981, 1988; Rubtzov, 1974; Peterson, 1977; Wood & Borkent, 1982; Currie, 1988). Proposed synapomorphies for Simuliidae exclusive of *Parasimulium* are provided by Wood & Borkent (1982) and Currie (1988). Significant support for this arrangement was observed in the following analyses: (1) 28S, all analyses (Table 4); (2) PEPCK, all codon weighting parsimony analyses of nucleotides and all

Table 5. Summary of support for relationships between Thaumaleidae and basal clades within Simuliidae as inferred from PEPCK. Numbers not in parentheses are bootstrap scores; numbers in parentheses are decay indices. 'NA' indicates no score was assigned to the node. '-' indicates the clade was present in fewer than 50% of the bootstrap replicates.

Locus/analysis PEPCK	Clades of interest				
	Thaumaleidae + Simuliidae	Simuliidae	Simuliinae	Prosimuliini <i>sensu</i> Currie (1988)	Simuliini <i>sensu</i> Currie (1988)
All taxa included					
DNA					
Parsimony					
1/1/1 codon weighting	92 (12)	- (0)	- (0)	98 (11)	- (0)
1/1/0 codon weighting	90 (6)	65 (2)	- (1)	- (2)	- (0)
3/5/1 codon weighting	96	65	51	96	70
3/10/1 codon weighting	87	58	74	90	81
3/20/1 codon weighting	84	58	68	80	81
Minimum evolution					
all positions	97	-	-	100	68
nt3 excluded	97	90	68	73	56
Neighbour-joining					
all positions	95	-	-	100	74
nt3 excluded	95	92	73	74	60
AMINO ACIDS					
Parsimony	-(0)	82 (3)	-(0)	-(0)	-(0)
Minimum evolution	62	98	80	82	73
Neighbour-joining	62	98	82	76	73
<i>Drosophila</i> excluded					
DNA					
Parsimony					
1/1/1 codon weighting	NA	-	-	98 (9)	67 (4)
1/1/0 codon weighting	NA	77 (3)	51 (1)	-(1)	-(0)
3/5/1 codon weighting	NA	70	74	93	76
3/10/1 codon weighting	NA	75	74	88	78
3/20/1 codon weighting	NA	72	69	75	77
Minimum evolution					
all positions	NA	61	72	100	90
nt3 excluded	NA	89	68	70	62
Neighbour-joining					
all positions	NA	51	63	100	93
100 nt3 excluded	NA	90	76	72	66
AMINO ACIDS					
Parsimony	NA	93 (5)	-(0)	-(0)	-(0)
Minimum evolution	NA	100	64	74	64
Neighbour-joining	NA	99	73	76	70

analyses of amino acids (Table 5); (3) DDC, 3/20/1 codon weighting parsimony analysis of nucleotides when *Drosophila melanogaster* is the distal outgroup taxon, all analyses of nucleotides, except NBJ, when *Aedes aegypti* is the distal outgroup, all analyses of nucleotides when *Cardiocladius* sp. and *Androsopoda gillespieae* are outgroup taxa and all analyses of amino acids, irrespective of outgroup taxon included (Table 6); (4) 12S, all analyses of both matrices (Table 7); (5) ND2, all analyses of nucleotides with third codon positions excluded and of amino acids (Table 7). Additionally, *Parasimulium crosskeyi* was more easily aligned with the thaumaleid than with the other simuliids in some portions of 28S, especially within hypervariable regions. This perhaps might be construed to

further attest to the considerable phylogenetic distance between *Parasimulium* and other simuliids.

Although attraction between the long branch of *Parasimulium crosskeyi* and those of the outgroup cannot be entirely dismissed, these genes, especially sites within 'conserved' regions of the 28S fragment examined and non-synonymous variation within PEPCK and DDC, do not appear to be evolving rapidly enough for it to be a problem. Because only four specimens belonging to the subgenus *Astoneomyia*, which represents the other side of the basal split within *Parasimulium*, are known, the long branch problem, if applicable, is unlikely to be resolved without including more representatives from related families.

Table 6. Summary of support for relationships between Chironomidae, Thaumaleidae and basal clades within Simuliidae as inferred from DDC. Numbers not in parentheses are bootstrap scores; numbers in parentheses are decay indices. 'NA' indicates no score was assigned to the node. '-' indicates the clade was present in fewer than 50% of the bootstrap replicates.

Locus/analysis DDC	Clades of interest					
	Chironomidae +Simuliidae	Simuliidae	Simuliinae	<i>Parasimulium</i> +Prosimuliini ¹	Prosimuliini <i>sensu</i> Currie (1988)	Simuliini <i>sensu</i> Currie (1988)
All taxa included						
DNA						
Parsimony						
1/1/1 codon weighting	– (0)	71 (4)	– (0)	72 (4)	87 (6)	– (0)
1/1/0 codon weighting	– (0)	84 (3)	– (0)	– (1)	88 (4)	53 (0)
3/5/1 codon weighting	–	85	–	58	94	53
3/10/1 codon weighting	–	88	–	–	91	74
3/20/1 codon weighting	–	90	58	–	83	92
Minimum evolution						
all positions	–	90	–	83	100	84
nt3 excluded	–	97	61	–	96	86
Neighbour-joining						
all positions	–	93	–	84	100	72
nt3 excluded	–	98	56	–	99	82
AMINO ACIDS						
Parsimony	– (0)	73 (0)	72 (2)	– (0)	87 (2)	60 (1)
Minimum evolution	–	92	86	–	99	93
Neighbour-joining	–	91	89	–	97	90
<i>Drosophila/Aedes</i> excluded						
DNA						
Parsimony						
1/1/1 codon weighting	– (0)/NA	97 (14)/100 (20)	56 (1)/77 (5)	– (0)/– (0)	96 (8)/93 (7)	90 (0)/92 (0)
1/1/0 codon weighting	– (0)/NA	82 (3)/98 (13)	– (1)/63 (2)	– (0)/– (0)	88 (4)/84 (3)	56 (0)/72 (0)
3/5/1 codon weighting	–/NA	93/100	56/91	–/–	95/97	86/96
3/10/1 codon weighting	50/NA	96/100	76/93	–/–	88/88	90/98
3/20/1 codon weighting	65/NA	94/100	77/96	–/–	84/78	98/98
Minimum evolution						
all positions	–/NA	98/99	66/91	–/–	100/100	98/98
nt3 excluded	–/NA	88/100	64/84	–/–	98/98	85/92
Neighbour-joining						
all positions	–/NA	100/100	–/70	55/–	100/100	96/98
nt3 excluded	–/NA	93/100	60/72	–/–	95/96	81/89
AMINO ACIDS						
Parsimony	– (0)/NA	64 (0)/98 (0)	74 (2)/80 (1)	– (0)/– (0)	96 (2)/92 (2)	54 (1)/71 (0)
Minimum evolution	–/NA	86/98	86/86	–/–	100/100	92/92
Neighbour-joining	–/NA	86/98	89/82	–/–	98/98	90/91

¹Prosimuliini *sensu* Currie (1988).

Parasimulium + *Prosimuliini sensu* Currie (1988) as sister group of all other Simuliidae. Stone (1963, 1965) considered *Parasimulium* to be most closely related to *Prosimulium* and allied genera and Borkent & Wood (1986) inferred this relationship using characters of the first-instar larva. The presumed synapomorphies used by Stone (1963, 1965) are unclear, although a widely separated fork in R_5 is one of the more likely candidates. The inference of Borkent & Wood (1986) was based largely upon characters of the larval mouthparts, many of which were described by Craig (1974) and Craig & Borkent (1980).

With *Drosophila* as the distant outgroup, unweighted, 1/1/0 and 3/5/1 codon weighting parsimony analyses and both distance analyses of all DDC nucleotides support this view (Fig. 2A, Table 6). Neighbour-joining analysis of DDC amino acids with *Aedes aegypti* as distant outgroup also recovers this topology with significant bootstrap support (Table 6). Support for this topology from DDC nucleotides diminishes with increased weighting of first and second positions and with the use of representatives from coninfraordinal families, rather than *Drosophila*, as outgroup. Increased weighting of first and second positions should be considered as making better use of

Table 7. Summary of support for relationships between Chironomidae, Thaumaleidae and basal clades within Simuliidae as inferred from 12S and ND2. Numbers not in parentheses are bootstrap scores; numbers in parentheses are decay indices. 'NA' indicates no score was assigned to the node. '-' indicate the clade was present in fewer than 50% of the bootstrap replicates.

Locus/analysis	Clades of interest						
	Thaumaleidae + Simuliidae	Simuliidae	Simuliinae	<i>Prosimulium s. str.</i> + Simuliinae ⁵	<i>Austrosimulium</i> + Prosimuliini ⁶	Prosimuliini <i>sensu</i> Currie (1988)	Simuliini <i>sensu</i> Currie (1988)
12S							
Parsimony ^{1/2}	– (0)/– (0)	70 (0)/83 (4)	67 (1)/55 (1)	– (0)/– (0)	– (0)/– (0)	– (1)/– (1)	– (1)/– (0)
Minimum evolution ^{1/2}	–/–	88/76	68/78	–/–	–/–	–/–	–/–
Neighbour-joining ^{1/2}	–/–	93/88	85/86	–/–	–/–	–/–	–/–
Parsimony ^{3/4}	56 (0)/NA	57 (0)/68 (3)	– (2)/74 (5)	– (0)/– (0)	– (0)/– (0)	– (1)/– (2)	– (1)/– (0)
Minimum evolution ^{3/4}	–/NA	72/85	70/70	–/–	–/–	–/–	–/–
Neighbour-joining ^{3/4}	–/NA	78/80	86/74	–/–	–/–	–/–	–/–
12S (Elision)							
Parsimony ^{1/2}	–/70	78/78	100/100	–/–	–/–	–/–	–/–
Minimum evolution ^{1/2}	63/84	100/100	100/95	–/–	–/–	61/84	92/–
Neighbour-joining ^{1/2}	74/98	100/100	100/100	–/–	–/–	61/60	84/88
Parsimony ^{3/4}	86/NA	79/99	100/100	–/73	–/–	–/–	–/–
Minimum evolution ^{3/4}	84/NA	100/94	90/100	–/–	–/–	60/–	78/56
Neighbour-joining ^{3/4}	88/NA	100/96	98/100	–/51	–/–	82/–	64/76
ND2							
DNA							
Parsimony							
1/1/1 codon weighting	NA	62 (4)	– (1)	– (0)	– (1)	– (2)	– (0)
1/1/0 codon weighting	NA	56 (1)	53 (0)	– (0)	– (0)	88 (3)	– (0)
3/5/1 codon weighting	NA	–	60	–	–	78	–
3/10/1 codon weighting	NA	–	58	–	–	77	–
3/20/1 codon weighting	NA	–	56	–	–	76	–
Minimum evolution							
all positions	NA	62	–	–	–	–	–
nt3 excluded	NA	60	78	–	–	82	51
Neighbour-joining							
all positions	NA	56	–	–	–	–	–
nt3 excluded	NA	64	78	–	–	80	–
AMINO ACIDS							
Parsimony	NA	76 (3)	61 (3)	– (0)	54 (2)	– (0)	– (0)
Minimum evolution	NA	96	78	–	–	81	–
Neighbour-joining	NA	94	81	–	–	79	–

¹All taxa included.²*Bittacomorpha* excluded.³*Anopheles* as distant outgroup.⁴*Androprosopa* and *Cardiocladius* as distant outgroup taxa.⁵Simuliinae *sensu* Crosskey (1988) and Currie (1988).⁶Prosimuliini *sensu* Currie (1988).

conserved variation in these sequences. *Drosophila melanogaster*, as a member of suborder Brachycera, is distantly related to simuliids in comparison to Culicidae, Chironomidae and Thaumaleidae and its use as an outgroup taxon may be problematical in that it probably represents a long branch (Felsenstein, 1978; Hillis, 1996).

Corroboration of a sister-group relationship between *Parasimulium* and all other simuliids from morphological data (Currie, 1988), analyses of this gene in which conserved variation is afforded higher weighting and results from most analyses of the other genes treated herein suggest that this topology is erroneous. Stone's (1965) placement of

Parasimulium may have been based upon symplesiomorphies. The inference of Borkent & Wood (1986) may have been seriously affected by the rooting of character states found in first-instar simuliid larvae with those found in mid- to late-instar larvae of outgroup families.

Tribal limits within Simuliinae. The following analyses significantly support division of Simuliidae, exclusive of *Parasimulium*, into monophyletic sister clades concordant with Grenier & Rageau's (1960) Prosimuliinae and Simuliinae and Currie's (1988) Prosimuliini and Simuliini: (1) 28S, all analyses, except minimum evolution analysis of data matrix 1 (Table 4); (2) EF-1 α , all analyses (Table 4); (3) PEPCK, all

analyses, except parsimony analyses of nucleotides using 1/1/1 and 1/1/0 codon weighting with all taxa included and 1/1/0 codon weighting with *Drosophila melanogaster* excluded and of amino acids (Table 5); (4) DDC, all analyses, except 1/1/1 codon weighting with all taxa included (Table 6); (5) 12S, minimum evolution and neighbour-joining analyses of the elision matrix, except for minimum evolution analysis with *Bittacomorpha clavipes* excluded (Table 7); (6) ND2, minimum evolution analysis of nucleotides with third positions excluded (Table 7).

Parsimony and neighbour-joining analyses of the 12S elision matrix with all outgroup taxa other than *Cardiocladius* sp. and *Androprosopa gillespieae* excluded and parsimony analysis of ND2 amino acids provide the only significant, albeit weak, support for a contradictory arrangement within Simuliinae (Fig. 2B, Table 7). These 12S elision analyses suggest *Prosimulium* may be the sister group to all remaining Simuliinae. A *Helodon* + *Urosimulium* + *Austrosimulium* clade was present in cladograms from several analyses of ND2 sequences, but only significantly (54% of the bootstrap replicates) in the parsimony analysis of ND2 amino acids. The weak support for these arrangements, limited taxon sampling, high A/T bias and lack of corroboration from morphological or other molecular evidence suggests both topologies are erroneous.

Summary of basal simuliid relationships. Basal relationships within Simuliidae now appear well established based upon concordance among six putatively independent data sets, one morphological (Currie, 1988) and five molecular (Moulton, 1997). Molecular evidence presented herein strongly supports a two-subfamily classification system for Simuliidae, *Parasimulium* s.l. as Parasimuliinae and all other simuliids as Simuliinae, and division of Simuliinae into monophyletic sister tribes, Prosimuliini and Simuliini (Fig. 1). Based upon complete congruence of these findings with Currie (1988) inference from morphological data, the tribal assignment of the seven simuliid genera that could not be included in this study is probably as follows: *Levitinia* is a prosimuliine and comprises *Araucnephia*, *Araucnephioides*, *Cnesiamima*, *Lutzsimulium*, *Sulcicnephia* and *Tlalocomyia* are simuliines. Therefore, Prosimuliini comprises *Helodon*, *Gymnopais*, *Levitinia*, *Prosimulium*, *Twinnia* and *Urosimulium*, and Simuliini comprises all remaining Simuliidae, except *Parasimulium*.

Among these genes, support for these relationships comes from a myriad of character types, ranging from synonymous (EF-1 α) and non-synonymous (particularly DDC and PEPCK) variation in protein-coding genes to substitutions in somewhat highly conserved portions of rDNA (28S) (Tables 4–6). Support from the latter class of characters is predicted upon high bootstrap and decay values for these relationships, even when variable regions are excluded (see Table 4). All observations in this study contradictory to the hypotheses immediately above are only weakly supported and all, except 12S and ND2, are supported entirely by third positions, which are affected by both saturation and nucleotide bias.

Through resolving the division of Simuliinae into monophyletic sister tribes, molecular evidence has, with

seemingly overwhelming support, elucidated a taxonomic problem that has plagued simuliid systematists for several decades. Several workers over the years have commented on the almost uninterrupted gradation or transformation of morphological characters from plesiomorphic to more apomorphic taxa, e.g. from *Prosimulium* and allies through '*Cnephia*' to *Simulium* (Dumbleton, 1963; Crosskey, 1969; Wygodzinsky & Coscarón, 1973), and some actually stated that no clear-cut basal division of the Simuliinae into two tribes, Prosimuliini and Simuliini, was possible using morphological characters. Out of the five synapomorphies Currie (1988) proposed for the monophyly of each tribe, two prosimuliine and three simuliine synapomorphies exhibit varying degrees of homoplasticity. Furthermore, two of the three remaining prosimuliine synapomorphies are sparsely surveyed and both of the remaining simuliine synapomorphies are regressive. Nonetheless, corroboration of his findings with molecular data presented here suggests these characters do exhibit phylogenetic utility and that his interpretation of them was impeccable.

Recognition of a clade containing *Prosimulium* (sometimes including *Helodon* and *Urosimulium*), *Gymnopais* and *Twinnia* has been commonplace (Shewell, 1958; Grenier & Rageau, 1960; Wood, 1978; Craig & Borkent, 1980; Borkent & Wood, 1986; Currie, 1988), whereas accurately identifying the relationship of this clade to other simuliid taxa has been a rarity. To my knowledge Grenier & Rageau (1960) and Currie (1988) are the only workers to have accomplished this feat. Grenier & Rageau's classification differs from that of Currie only in the relative ranking of taxa and the number of genera recognized. All other classification schemes should be rejected because they contain unnatural groupings. Those not reflecting a sister-group relationship between *Parasimulium* and the remainder of Simuliidae and those affording a clade containing *Gymnopais*, *Levitinia* and *Twinnia* a higher taxonomic ranking than one containing *Prosimulium*, *Helodon* and *Urosimulium*, e.g. Rubtzov (1974) and Py-Daniel & Moreira Sampaio (1994), are particularly untenable.

The stage has now been set to resolve intratribal relationships, particularly within Simuliini. This tribe represents the bulk of generic and species diversity within Simuliidae and resolution of relationships within it is critical in order to understand more fully character evolution and biogeographical patterns within Simuliidae and the origin of the pest genus *Simulium*. Relationships within Simuliini and, to a lesser extent, Prosimuliini, as inferred from sequences presented herein, will be the focus of a forthcoming paper.

Acknowledgements

D. R. Maddison, M. F. Hammer, L. A. McDade, R. E. Smith and M. F. G. Taylor provided guidance in this course of this research. D. R. Maddison provided keen insights, laboratory space and reagents, and numerous letters of recommendation that played a large role in obtaining financial support for this research. J. Prince (University of Western Australia) and P. Cranston (CSIRO, Canberra) were gracious hosts while I was

in Australia collecting simuliids. D. R. Maddison, K. P. Pruess and D. G. Mead reviewed earlier drafts of this manuscript and their comments improved it significantly. The following people kindly supplied material used in this study: P. H. Adler (Clemson University), D. B. Bedo (DPI, Queensland), P. Cranston (CSIRO, Canberra), D. C. Currie (University of Toronto), D. M. Wood (ECORC, retired), R. W. Crosskey (NHM, retired), S. Coscarón (University of La Plata), R. Miranda Esquivel (University of La Plata), F. C. de Moor (Albany Museum, Grahamstown, South Africa) and B. E. Marshall (University of Zimbabwe). This research was supported in the most part by National Science Foundation Doctoral Dissertation Improvement Grant DEB-9520613 awarded to J.K.M. Additional support came from the Flinn Foundation, the NSF Research Training Group (RTG) Grant in the Analysis of Biological Diversification, DIR-9113362, awarded to the University of Arizona, the Graduate College of the University of Arizona and NSF grant DEB-9420219 awarded to D.R.M.

References

- Anderson, S., DeBruijn, M., Coulson, A., Eperon, I., Sanger, F. & Young, I. (1982) Complete sequence of the bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *Journal of Molecular Biology*, **156**, 683–717.
- Bakker, F.T., Olsen, J.L., Stam, W.T. & Van Den Hoek, C. (1994) The *Cladophora* complex (Chlorophyta): new views based upon 18S rRNA gene sequences. *Molecular Phylogenetics and Evolution*, **3**, 365–382.
- Ballard, J.W.O. (1994) Sequence data resolves a morphological conundrum in *Austrosimulium* (Diptera: Simuliidae). *Journal of the Australian Entomological Society*, **33**, 131–135.
- Ballard, J.W.O., Olsen, G.J., Faith, D.P., Odgers, W.A., Rowell, D.M. & Atkinson, W.P. (1992) Evidence from 12S ribosomal RNA sequences suggest that onychophorans are modified arthropods. *Science*, **258**, 1345–1348.
- Beard, C.B., Mills Hamm, D. & Collins, F.H. (1993) The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences from other insects. *Insect Molecular Biology*, **2**, 103–124.
- Borkent, A. & Wood, D.M. (1986) The first and second larval instars and the egg of *Parasimulium stonei* Peterson. *Proceedings of the Entomological Society of Washington*, **88**, 287–296.
- Bremer, K. (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution*, **42**, 795–803.
- Brockhouse, C.L., Vajime, C.G., Marin, R. & Tanguay, R.M. (1993) Molecular identification of onchocerciasis vector sibling species in black flies (Diptera: Simuliidae). *Biochemical and Biophysical Research Communications*, **194**, 628–634.
- Bruns, T.D., Vilgalys, R., Barns, S.M., Gonzalez, D., Hibbett, D.S., Lane, D.J., Simon, L., Stickel, S. & Szaro, T.M. (1992) Evolutionary relationships within the Fungi: analyses of nuclear small subunit rRNA sequences. *Molecular Phylogenetics and Evolution*, **1**, 231–241.
- Cho, S., Mitchell, A., Regier, J.C., Mitter, C., Poole, R.W., Friedlander, T.P. & Zhao, S. (1995) A highly conserved nuclear gene for low-level phylogenetics: *elongation factor-1-a* recovers morphology-based cladogram for heliothine moths. *Molecular Biology and Evolution*, **12**, 650–656.
- Clary, D.O. & Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, **22**, 252–271.
- Collins, T.M., Kraus, F. & Estabrook, G. (1994) Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. *Systematic Biology*, **43**, 449–459.
- Coscarón, S. (1985) On *Cnesiamima atroparva* Edwards larva and its significance to interpret the phylogenetic relationships of notogeic black flies (Insecta, Diptera, Simuliidae). *Boletín of the Sociedad de Biología de Concepcion*, **56**, 67–70.
- Craig, D.A. (1974) The labrum and cephalic fans of Simuliidae (Diptera: Nematocera). *Canadian Journal of Zoology*, **52**, 133–159.
- Craig, D.A. (1977) A re-assessment of the systematic position of *Pseudosimulium humidum* (Westwood), an Upper Jurassic fossil dipteran. *Quaestiones Entomologicae*, **13**, 175–179.
- Craig, D.A. & Borkent, A. (1980) Intra- and interfamilial homologies of maxillary palpal sensilla of larval Simuliidae (Diptera: Culicomorpha). *Canadian Journal of Zoology*, **58**, 2264–2279.
- Crosskey, R.W. (1969) A re-classification of the Simuliidae (Diptera) of Africa and its islands. *Bulletin of the British Museum of Natural History (Entomology), Supplement*, **14**, 1–195.
- Crosskey, R.W. (1981) Simuliid taxonomy – the contemporary scene, *Blackflies: the Future for Biological Methods in Integrated Control* (ed. by M. Laird), pp. 3–18. Academic Press, London.
- Crosskey, R.W. (1988) An annotated checklist of the world's black flies (Diptera: Simuliidae), *Black Flies: Ecology, Population Management and Annotated World List* (ed. by K. C. Kim and R. W. Merritt), pp. 425–520. Pennsylvania State Press, University Park, Pennsylvania.
- Crosskey, R.W. (1991) The fossil pupa *Simulimima* and the evidence it provides for the Jurassic origin of the Simuliidae (Diptera). *Systematic Entomology*, **16**, 401–406.
- Crosskey, R.W. & Howard, T.M. (1997) *A New Taxonomic and Geographical Inventory of World Blackflies (Diptera: Simuliidae)*. The Natural History Museum, London.
- Crozier, R.H. & Crozier, Y.C. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, **133**, 97–117.
- Currie, D.C. (1988) *Phylogeny of Primitive Simuliidae (Insecta: Diptera: Culicomorpha)*. PhD Dissertation, University of Alberta, Edmonton, Canada.
- Dixon, M.T. & Hillis, D.M. (1991) Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. *Molecular Biology and Evolution*, **10**, 256–267.
- Dumbleton, L.J. (1963) The classification and distribution of the Simuliidae (Diptera) with particular reference to the genus *Austrosimulium*. *New Zealand Journal of Science*, **6**, 320–357.
- Edwards, F.W. (1934) The Simuliidae (Diptera) of Java and Sumatra. *Archives für Hydrobiologie*, **13**, 92–138.
- Enderlein, G. (1937) Über bemerkenswerte Unkenntnisse von der Entwicklung zweier einheimischer Simuliiden Gattungen (Diptera). *Markische Tierwelt*, **3**, 68–70.
- Fang, Q.Q., Cho, S., Regier, J.C., Mitter, C., Matthews, M., Poole, R.W., Friedlander, T.P. & Zhao, S. (1997) A new nuclear gene for insect phylogenetics: dopa decarboxylase is informative of relationships within Heliothinae (Lepidoptera: Noctuidae). *Systematic Biology*, **46**, 269–283.
- Felsenstein, J. (1978) Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology*, **27**, 401–410.
- Flook, P.K., Rowell, C.H.F. & Gellissen, G. (1995) The sequence,

- organization, and evolution of the *Locusta migratoria* mitochondrial genome. *Journal of Molecular Evolution*, **41**, 928–941.
- Friedlander, T.P., Regier, J.C. & Mitter, C. (1992) Nuclear gene sequences for higher-level phylogenetic analysis: 14 promising candidates. *Systematic Biology*, **41**, 483–489.
- Friedlander, T.P., Regier, J.C., Mitter, C. & Wagner, D.L. (1996) A nuclear gene for higher level phylogenetics: phosphoenolpyruvate carboxykinase tracks Mesozoic-age divergences within Lepidoptera (Insecta). *Molecular Biology and Evolution*, **13**, 594–604.
- Friedrich, M. & Tautz, D. (1997) Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. *Systematic Biology*, **46**, 674–698.
- Gilbert, D.G. (1992) *SeqApp, Version 1.9a157*. Available via anonymous FTP: ftp.bio.indiana.edu, in folder /molbio/seqapp, as seqapp.hqx.
- Grenier, P. & Rageau, J. (1960) Simulies (Dipt., Simuliidae) di Tahiti. Remarques sur la classification des Simuliidae. *Bulletin of the Societe de Pathologie Exotique*, **53**, 727–742.
- Hasegawa, M. & Hashimoto, T. (1993) Ribosomal RNA cladograms misleading? *Nature (London)*, **361**, 23.
- Hasegawa, M., Kishino, H. & Yano, T. (1985) Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **21**, 160–174.
- Hedges, S.B., Parker, P.H., Sibley, C.S. & Kumar, S. (1996) Continental breakup and the ordinal diversification of birds. *Nature (London)*, **381**, 226–229.
- Hennig, W. (1973) Ordnung Diptera (Zweiflügler). *Handbuch der Zoologie, Berlin*, **4**(2), 1–337.
- Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J. & Penny, D. (1996) Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Molecular Biology and Evolution*, **13**, 150–169.
- Hillis, D.M. (1996) Inferring complex phylogenies. *Nature (London)*, **383**, 130–131.
- Hoc, L.V.L., Lecointre, G. & Perasso, R. (1993) A 28S rRNA-based phylogeny of the gnathostomes: first steps in the analysis of conflict and congruence with morphologically based cladograms. *Molecular Phylogenetics and Evolution*, **2**, 31–51.
- Jell, P.A. & Duncan, P.M. (1986) Invertebrates, mainly insects, from the freshwater, Lower Cretaceous, Koonwarra fossil bed (Korumburra Group), South Gippsland, Victoria. *Memoirs of the Association of Australasian Paleontologists*, **3**, 111–205.
- Kalugina, N.S. (1991) New Mesozoic Simuliidae and Leptoconopidae and the origin of blood-sucking in lower dipterans. *Paleontologicheskii Zhurnal*, **1**, 69–80.
- Kalugina, N.S. & Kovalev, V.G. (1985) *Jurassic Diptera of Siberia*. Nauka, Moscow.
- Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X. & Wilson, A.C. (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science U.S.A.*, **86**, 6196–6200.
- Kraus, F., Jarecki, L., Miyamoto, M.M., Tanhouser, S.M. & Paipis, P.J. (1992) Mispairing and compensatory changes during the evolution of mitochondrial ribosomal RNA. *Molecular Biology and Evolution*, **9**, 770–774.
- Madahar, D.P. (1969) The salivary gland chromosomes of seven taxa in the subgenus *Stegopterna* (Diptera, Simuliidae, *Cnephia*). *Canadian Journal of Zoology*, **47**, 115–119.
- Maddison, W.P. & Maddison, D.R. (1992) *MacClade: Analysis of Phylogeny and Character Evolution, Version 3.0*. Sinauer Associates, Sunderland, Massachusetts.
- Miller, B.R., Crabtree, M.B. & Savage, H.M. (1997) Phylogenetic relationships of the Culicomorpha inferred from 18S and 5.8S ribosomal DNA sequences (Diptera: Nematocera). *Insect Molecular Biology*, **6**, 105–114.
- Mitchell, A., Cho, S., Regier, J.C., Mitter, C., Poole, R.W. & Matthews, M. (1997) Phylogenetic utility of *elongation factor-1 α* in Noctuoidea (Insecta: Lepidoptera): the limits of synonymous substitution. *Molecular Biology and Evolution*, **14**, 381–390.
- Mitchell, S.E., Cockburn, A.F. & Seawright, J.A. (1993) The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*, **36**, 1058–1073.
- Miyamoto, M.M., Allard, M.W., Adkins, R.M., Janecek, L.L. & Honeycutt, R.L. (1994) A congruence test of reliability using linked mitochondrial DNA sequences. *Systematic Biology*, **43**, 236–249.
- Moulton, J.K. (1997) *Molecular Systematics of the Simuliidae (Diptera: Culicomorpha)*. PhD Dissertation, University of Arizona, Tucson, Arizona.
- Oosterbroek, P. & Courtney, G.W. (1995) Phylogeny of the nematoceros families of Diptera (Insecta). *Zoological Journal of the Linnean Society*, **115**, 267–311.
- Pawlowski, J., Szadziewski, R., Kmiecik, D., Fahrni, J. & Bittar, G. (1996) Phylogeny of the infraorder Culicomorpha (Diptera: Simuliidae) based upon 28S RNA gene sequences. *Systematic Entomology*, **21**, 167–178.
- Pélandakis, M. & Solignac, M. (1993) Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. *Journal of Molecular Evolution*, **37**, 525–543.
- Perasso, R. & Baroin Tourancheau, A. (1993) The eukaryogenesis A model derived from rRNA phylogenies. *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales*, **186**, 656–665.
- Peterson, B.V. (1977) A synopsis of the genus *Parasimulium* Malloch (Diptera: Simuliidae), with descriptions of one new subgenus and two new species. *Proceedings of the Entomological Society of Washington*, **79**, 96–106.
- Peterson, B.V. (1989) An unusual black fly (Diptera: Simuliidae), representing a new genus and new species. *Journal of the New York Entomological Society*, **97**, 317–331.
- Post, R.J. & Flook, P. (1992) DNA probes for the identification of members of the *Simulium damnosum* complex (Diptera: Simuliidae). *Medical and Veterinary Entomology*, **6**, 379–384.
- Pruess, K.P., Zhu, X. & Powers, T.O. (1992) Mitochondrial transfer RNA genes in a black fly, *Simulium vittatum* (Diptera: Simuliidae), indicate long divergence from mosquito (Diptera: Culicidae) and fruit fly (Diptera: Drosophilidae). *Journal of Medical Entomology*, **29**, 644–651.
- Py-Daniel, V. & Moreira Sampaio, R.T. (1994) *Jalacinglomyia* gen. n. (Culicomorpha); a ressurreicao de Gymnopauidinae; a eliminacao do nivel tribal; apresentacao de novos caracteres e a redescricao dos est-ges larval y pupal de *Simulium colombaschense* (Fabricius, 1787) (Diptera: Simuliidae). *Memorias de CAICET*, **4**, 101–148.
- Rohdendorf, B. (1964) The historical development of Diptera. *Trudy Paleontologicheskogo Instituta*, **100**, 1–311.
- Rothfels, K. (1979) Cytotaxonomy of black flies (Simuliidae). *Annual Review of Entomology*, **24**, 507–539.
- Rothfels, K. (1981) Cytotaxonomy: principles and their application to some northern species-complexes in *Simulium*. *Blackflies: the Future for Biological Methods in Integrated Control* (ed. by M. Laird), pp. 19–29. Academic Press, New York.
- Rubtzov, I.A. (1956) Fauna USSR. Insects Diptera. Black Flies (Family Simuliidae). *Zoological Institute of the Academy of Sciences of the USSR (New Series, 64)*, **6**(6), 1–860.
- Rubtzov, I.A. (1974) *Evolution, Phylogeny, and Classification of the*

- Family Simuliidae (Diptera) [translated from Russian]. British Library Lending Division, Wetherby.
- Shewell, G.E. (1958) Classification and distribution of arctic and subarctic Simuliidae. *Proceedings of the 10th International Congress of Entomology*, **1**, 635–643.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.
- Smart, J. (1945) The classification of the Simuliidae (Diptera). *Transactions of the Royal Society of London*, **95**, 463–532.
- Smith, A.B. (1989) RNA sequence data in phylogenetic reconstruction: testing the limits of resolution. *Cladistics*, **5**, 321–344.
- Steel, M.A., Lochart, P.J. & Penny, D. (1993) Confidence in evolutionary cladograms from biological sequence data. *Nature (London)*, **364**, 440–442.
- Stone, A. (1963) *An Annotated List of Genus-Group Names in the Family Simuliidae (Diptera)*. United States Department of Agriculture, Technical Bulletin 1284. United States Department of Agriculture, Washington, DC.
- Stone, A. (1965) Family Simuliidae. *A Catalogue of the Diptera of America North of Mexico* (ed. by A. Stone), pp. 181–189. Agriculture Handbook no. 276. United States Department of Agriculture, Washington, DC.
- Swofford, D.L. (1996–97) *PAUP**, Versions 4.0d52, 4.0d54 & 4.0d56 (beta test versions, used with Permission of D. L. Swofford; completed version 4.0 to be distributed by Sinauer Associates, Sunderland, Massachusetts).
- Tang, J., Toe, L., Back, C., Zimmerman, P.A., Pruess, K. & Unnasch, T.R. (1995) The *Simulium damnosum* species complex: phylogenetic analysis and molecular identification based upon mitochondrially encoded gene sequences. *Insect Molecular Biology*, **4**, 79–88.
- Tang, J., Toe, L., Pruess, K., Cupp, E.W. & Unnasch, T.R. (1996) Molecular phylogeny and typing of blackflies (Diptera: Simuliidae) that serve as vectors of human or bovine onchocerciasis. *Medical and Veterinary Entomology*, **10**, 228–234.
- Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1996) *Clustal W, Version 1.6*. Available at the <http://biomaster.uio.no/clustalw.html>.
- Vossbrinck, C.R. (1989) A 28S ribosomal RNA phylogeny of certain cyclorrhaphous Diptera based upon a hypervariable region. *Systematic Entomology*, **14**, 417–431.
- Walldorf, U., Hovemann, B. & Bautz, E.K.F. (1985) F1 and F2: two similar genes regulated differently during development of *Drosophila melanogaster*. *Proceedings of the National Academy of Science U.S.A.*, **82**, 5795–5799.
- Wheeler, W.C., Gatesy, J. & DeSalle, R. (1995) Elision: a method for accommodating multiple molecular sequence alignments with alignment-ambiguous sites. *Molecular Phylogenetics and Evolution*, **4**, 1–9.
- Wheeler, W.C. & Honeycutt, R.L. (1988) Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Molecular Biology and Evolution*, **5**, 90–96.
- Wood, D.M. (1978) Taxonomy of the Nearctic species of *Twinnia* and *Gynnopais* (Diptera: Simuliidae) and a discussion of the ancestry of the Simuliidae. *Canadian Entomologist*, **110**, 1297–1337.
- Wood, D.M. & Borkent, A. (1982) Description of the female of *Parasimulium crosskeyi* Peterson (Diptera: Simuliidae), and a discussion of the phylogenetic position of the genus. *Memoirs of the Entomological Society of Washington*, **10**, 193–210.
- Wood, D.M. & Borkent, A. (1989) Phylogeny and classification of the Nematocera. *Manual of Nearctic Diptera*, Vol. 3 (ed. by J. F. McAlpine and D. M. Wood), pp. 1333–1370. Research Branch Agriculture Canada Monograph 32. Ottawa.
- Wygodzinsky, P. & Coscarón, S. (1973) A review of the Mesoamerican and South American black flies of the tribe Prosimuliini (Simuliinae, Simuliidae). *Bulletin of the American Museum of Natural History*, **151**, 129–200.
- Wygodzinsky, P. & Díaz Nájera, A. (1970) Un nuevo genero de Simulido de la Republica Mexicana (Diptera: Simuliidae). *Revista de Investigacion en Salud Publica*, **30**, 83–110.
- Xiong, B. & Kocher, T.D. (1991) Comparison of mitochondrial sequences of seven morphospecies of black flies (Diptera: Simuliidae). *Genome*, **34**, 306–311.
- Xiong, B. & Kocher, T.D. (1993) Intraspecific variation in sibling species of *Simulium venustum* and *Simulium verecundum* complexes (Diptera: Simuliidae) revealed by the sequence of the mitochondrial 16S rRNA gene. *Canadian Journal of Zoology*, **71**, 1202–1206.
- Yankovsky, A.V. (1992) A review of the system of blackflies (Diptera, Simuliidae) on the suprageneric level. *Parazitologicheskyy Sbornik*, **37**, 203–217 (in Russian, with English summary).
- Zuker, M. & Jaeger, J. (1990) *Mulfold, Version 2.0d82*. Available via anonymous FTP: <ftp://ftp.bio.indiana.edu>, in folder /molbio/, as mulfold.hqx.

Accepted 8 March 1999