Molecular Phylogenetics and Evolution
Vol. 16, No. 2, August, pp. 286–295, 2000
doi:10.1006/mpev.2000.0807, available online at http://www.idealibrary.com on IDEAL®

Utility of the Mitochondrial Cytochrome Oxidase II Gene for Resolving Relationships among Black Flies (Diptera: Simuliidae)

K. P. Pruess,* B. J. Adams,†,‡ T. J. Parsons,†,§ X. Zhu,* and T. O. Powers†

*Department of Entomology, University of Nebraska, Lincoln, Nebraska 68583-0816; †Department of Plant Pathology, University of Nebraska, Lincoln, Nebraska 68583; ‡Department of Nematology, University of Florida, Gainesville, Florida 32601; and §Armed Forces DNA Identification Laboratory, Rockville, Maryland 20850

Received August 31, 1999, revised May 3, 2000

The complete mitochondrial cytochrome oxidase II gene was sequenced from 17 black flies, representing 13 putative species, and used to infer phylogenetic relationships. A midge (*Paratanytarsus* sp.) and three mosquitoes (*Aedes aegypti, Anopheles quadrimaculatus,* and *Culex quinquefasciatus*) were used as outgroup taxa. All outgroup taxa were highly divergent from black flies. Phylogenetic trees based on weighted parsimony (a priori and a posteriori), maximum likelihood, and neighbor-joining (log-determinant distances) differed topologically, with deeper nodes being the least well-supported. All analyses supported current classification into species groups but relationships among those groups were poorly resolved. The majority of phylogenetic signal came from closely related sister taxa. The CO-II gene may be useful for exploring relationships at or below the subgeneric level, but is of questionable value at higher taxonomic levels. The weighting method employed gave phylogenetic results similar to those reported by other authors for other insect CO-II data sets. A best estimate of phylogenetic relationships based on the CO-II gene is presented and discussed in relation to current black fly classification.

Key Words: **black fly; Simuliidae; mitochondrial DNA; cytochrome oxidase subunit II; phylogeny; character weighting.**

INTRODUCTION

Black flies (Diptera: Simuliidae) are a large group of medically and economically important Diptera in the suborder Nematocera which have speciated with little morphological change. A paucity of diagnostic characters for all life stages complicates species identification and therefore control efforts often target black flies in general, rather than the few pest species.

Previous workers agree that Simuliidae is a monophyletic group. Wood and Borkent (1989) and Oosterbroek (1995) place Simuliidae in the superfamily Chironomoidea with Chironomidae (midges) and Ceratopogonidae (biting midges) as sister groups. These taxa are included, with Culicidae (mosquitoes) and other families, in the infraorder Culicomorpha. Molecular data, based on 28s rDNA gene sequences (Pawlowski *et al.*, 1996), did not support these relationships, suggesting that Thaumaleidae (solitary midges) and Simuliidae are sister groups and that Chironomidae may be a sister group to all other Nematocera. Using a different 28s rDNA sequence, Friedrich and Tautz (1997) found the Culicomorpha to form a strongly supported monophyletic clade but the relationships between families were poorly resolved.

The fossil record for Simuliidae is meager. The earliest known fossil, a pupa from middle Jurassic (170 mybp) is considered indistinguishable from modern *Prosimulium*, and larvae from the lower Cretaceous (120 mybp) of Australia are indistinguishable from modern *Simulium* (Crosskey, 1990). Quite likely, the divergence of modern genera was complete in the Triassic.

Two subfamilies are commonly recognized in Simuliidae, with all but four species placed in the subfamily Simuliinae. Simuliinae is divided into two tribes, Prosimuliini and Simuliini, but with disagreement as to tribal placement of some genera. The genus *Simulium* contains over 1000 species (Crosskey and Howard, 1997). The current classification of black flies recognizes "species groups" within subgenera without inferring relationships among those groups (Crosskey and Howard, 1997). Currie (1988) provides an alternative classification at the generic level.

Black fly taxonomy is further complicated by sibling species complexes whose component species cannot be morphologically differentiated in all life stages. Many sibling species are best characterized by larval salivary gland chromosome polymorphisms. Sibling species are inferred from lack of heterozygotes in areas of sympatry (Rothfels, 1979). Larvae of these sibling species are usually isolated temporally, spatially, or ecologically (Adler, 1987). Within species groups, a number of cy-

1055-7903/00 $35.00

TABLE 1

Collection Data for Black Flies Sequenced

Species	Source
Heledon dicentum Dyar & Shannon	California, Napa Co., 22 March 1990
Prosimulium onychodactylum Dyar & Shannon	Colorado, Jackson Co., 5 Aug. 1989
Cnephia sp. A (putative *C. dacotensis* Dyar & Shannon)	New Hampshire, Strafford Co., 4 May 1990
C. sp. B [putative *C. pecuarum* (Riley)]	Texas, Rains Co., 4 Jan. 1991
C. sp. C (putative *C. pecuarum*)	Texas, Rains Co., 4 Jan. 1991
Simulium (Eusimulium) pilosum (Knowlton & Rowe) (= *S. aureum* "B")	Nebraska, Lancaster Co., 24 April 1989
S. (Eusimulium) donovani Vargas G1 (= *S. aureum* "G")	California, Mono Co., 25 June 1991
S. (Eusimulium) donovani G2	California, Mono Co., 25 June 1991
S. (Psilozia) vittatum Zetterstedt sibling IIIL-1	Nebraska, Lancaster Co., 25 March 1988
S. (Psilozia) vittatum "Iceland"	Iceland, R. Laxa, 31 Aug. 1990
S. (Psilozia) encisoi Vargas & Diaz Najera	New Mexico, Dona Ana Co., 2 June 1992
S. (Psilopelmia) bivittatum Malloch A	Nebraska, Loup Co., 12 Oct. 1990
S. (Psilopelmia) bivittatum B	Wyoming, Crook Co., 21 May 1994
S. (Simulium) piperi Dyar & Shannon (*hunteri* group)	Nebraska, Scottsbluff Co., 11 July 1989
S. (Simulium) jacumbae Dyar & Shannon (*hunteri* group)	Nebraska, Lancaster Co., 30 Oct. 1992
S. (Simulium) luggeri Malloch (*jenningsi* group)	Nebraska, Loup Co., 12 July 1990
S. (Simulium) confusum Moulton & Adler (*jenningsi* group)	Nebraska, Saline Co., 6 July 1990
Outgroups	
Paratanytarsus sp. (Chironomidae)	Pennsylvania, Chester Co., Jan. 1993
Aedes aegypti (L.) (Culicidae)	Ho *et al.* (1995) (GenBank L34412)
Anopheles quadrimaculata Say (Culicidae)	Mitchell *et al.* (1993) (GenBank L04272)
Culex quinquefasciatus Say (Culicidae)	Ho *et al.* (1995) (GenBank L34351)

tophylogenies based on chromosomal inversions exist (e.g., Rothfels, 1979), but these are unrooted.

The cytochrome oxidase subunit II (CO-II) gene is one of the best-known mitochondrial genes. Cooper *et al.* (1991) and Saraste (1990) review the structure and function of the polypeptide subunits. CO-II contains both highly conserved and variable regions. The sequence of this gene is potentially useful for phylogenetic analyses over a wide taxonomic range and has been explored for that purpose. However, Liu and Beckenbach (1992) were unable to resolve the evolution of insect orders and concluded that divergence may have occurred too rapidly for resolution using mitochondrial genes. This sequence has been generally useful for reconstructing phylogenies among more closely related groups (Willis *et al.*, 1992; Beckenbach *et al.*, 1993; Brown *et al.*, 1994; Sperling and Hickey, 1994; Emerson and Wallis, 1995; Spicer, 1995; Smith and Bush, 1997; Foley *et al.*, 1998). With the exception of the Collembola (Frati *et al.*, 1997), the utility of this sequence has not been well explored at the generic or lower levels for insects.

We obtained the complete sequence of the mitochondrial CO-II gene for 17 black flies. Our primary purpose was to explore the utility of this gene for resolving relationships among black flies at different taxonomic levels (tribe to sibling species). A phylogeny based on parsimony analysis, using empirically derived weights, is presented and discussed in relation to current black fly classification.

MATERIALS AND METHODS

Black Fly Collection and DNA Extraction

Larval black flies were either preserved in ethanol upon collection or returned to the laboratory alive and preserved individually in TE buffer at −70°C. The midge was obtained as an adult. With two exceptions (to be discussed), all specimens were obtained from localities where, according to collecting and rearing data, other closely related species did not occur. Collection data are given in Table 1. DNA was obtained using proteinase K digestion, phenol/chloroform extraction, and ethanol precipitation as described by Wilkerson *et al.* (1993).

PCR and Sequencing

Using the notation of Simon *et al.* (1994), primers used to amplify the complete CO-II gene were CI-J-2797, 5′-CCTCGACGTTATTCAGATTACC-3′; C2-J-3123, 5′-GGACTACAAGATAGAGCCTC-3′; C2-J-3407, 5′-CATCAATGATACTGAAGTTATGA-3′; C2-J-3684, 5′-GGTCAATGTTCAGAAATTTGTGG-3′; C2-N-3380, 5′-TCAATATCATTGATGACCAAT-3′; C2-N-3686, 5′-CAATTGGTATAAAACTATGATTTGC-3′; A8-N-3926, 5′-AATAAAGATAATCAACTAATAGC-3′; and A6-N-4126, 5′-AGTCCTAAGAATGTTCTTA-ATCA-3′.

When possible, the entire CO-II gene was amplified using primers C1-J-2797 and A6-N-4126 or A8-N-3926

with other internal primers used for sequencing. In a few cases, other overlapping combinations were employed.

Two sequencing methods were used. Most sequence data were obtained by direct sequencing of double-stranded PCR products. Primers were end-labeled with T4 polynucleotide kinase (Promega, Madison, WI) and γ-^{32}P and used with Promega's fmol sequencing system according to manufacturer's recommendations. Alternatively, products were cloned before sequencing. Amplified products were purified with Gene Clean II (Bio 101, La Jolla, CA), end-repaired with T4 DNA Polymerase (Promega), ethanol precipitated, and ligated into *Sma*cut pBlue-Script SK+ vector (Stratagene, La Jolla, CA). Ligated DNAs were transformed into Stratagene *Escherichia coli* XL1-Blue supercompetent cells and plated on selective medium with blue/white color selection. Individual white colonies were screened for proper size insert using the initial PCR amplification conditions except the number of cycles was reduced to 23. Colonies with appropriately sized inserts were grown, and single-strand DNA was recovered using VCS M13 Helper Phage (Stratagene). Sequences were obtained using Sequenase kit version 2.0 (U.S. Biochemical, Cleveland, OH) with ^{35}S-labeling.

To the extent possible, both strands were sequenced. Sequence of the opposite strand usually resolved problems due to compressions in direct sequence. Multiple clones (3 or 4) were sequenced for each taxon to determine a consensus; a few instances of presumed *Taq* polymerase error were observed in single clones. Sequences were derived from a single individual of most species. In two cases, however, additional individuals were sequenced to help resolve questions about the species involved. We include all sequences that differ. Sequences reported in this paper have been submitted to the GenBank DNA sequence bank and are available under Accession Nos. M76433 and AF08357–AF08373.

Phylogenetic Analysis

Published sequences for three mosquitoes, *Aedes aegypti* (Ho *et al.*, 1995), *Anopheles quadrimaculata* (Mitchell *et al.*, 1993), and *Culex quinquefasciatus* (Ho *et al.*, 1995), were included as outgroup taxa. We also sequenced a chironomid, *Paratanytarsus* sp., presumed to be more closely related to black flies. DNA sequences were aligned by eye. We also analyzed the data using *Paratanytarsus* as the sole outgroup taxon. Phylogenetic signal was estimated using relative apparent synapomorphy analysis (RASA 2.3) (Lyons-Weiler, 1999; Lyons-Weiler *et al.*, 1996). To identify the effect of taxon sampling on signal content, we did separate analyses on pruned sets of taxa (i.e., outgroup taxa, closely related sister taxa removed). Cunningham (1997) showed that six-parameter weighted parsimony can give consistent positive results between congruence and accuracy and in the present study weights were derived a priori based on a modification of the six-parameter model of Rodriquez *et al.* (1990) (Appendix). This model assumes that base frequencies of all sequences, including the ancestral sequence, are at equilibrium. All possible pairwise comparisons of nucleotide use were made by codon position. We evaluated weights based on nucleotide use at only variable sites and also at all sites (Waddell and Steel, 1997). Outgroup taxa were included in computing weights except for the last six codons where alignment was questionable. The weights used were proportional to the ratio of expected/observed values based on nucleotide frequency. Integer values were entered into step matrices for analysis with PAUP*4.0.0d64 for PPC and UNIX (provided by David L. Swofford). Starting trees for heuristic searches using branchswapping were obtained by stepwise addition. Branchswapping options consisted of random addition sequence using 30 replications with tree bisection–reconnection (TBR).

A heuristic search was also used to recover a maximum likelihood (ML) tree. Search options included getting the starting tree by stepwise addition, using the "as-is" algorithm of sequence addition, and TBR branchswapping. The substitution model for the maximum likelihood analysis assumed unequal transition-to-transversion rates. The ratio of transitions to transversions and nucleotide base frequencies were estimated from the data set. The analysis also assumed the HKY two-parameter model for unequal base frequencies (Hasegawa *et al.*, 1985). Variable sites were assumed to reflect a gamma distribution with six rate categories, the shape parameter being estimated from the data set. Additionally, a neighbor-joining tree was produced using log-determinant distances.

Competing topologies were evaluated using parsimony and maximum likelihood criteria according to Kishino and Hasegawa (1989) and Templeton's nonparametric test of parsimony (Templeton, 1983) using the Treescores option in PAUP*. To study the effect of the outgroup taxa on tree structure, we removed all of the outgroup taxa from the analysis with the exception of *Paratanytarsus* and duplicated the tree-building analyses. Because the relationships among the outgroup taxa could alter tree lengths and likelihoods of the overall tree, and because some trees were generated in the absence of these taxa, all outgroup taxa (except *Paratanytarsus*) were removed for tree comparisons. For parsimony trees, nodal support was estimated by bootstrapping (500 replicates), and Bremer support indices (Bremer, 1988, 1994; Kallersjo *et al.*, 1992) were calculated using AutoDecay version 2.9.7 (Eriksson, 1997). Trees were visualized using TreeView version 1.5 (Page, 1998).

TABLE 2

Nucleotide Use by Codon Position in Black Flies

Position	Nucleotide A	C	G	T
1	1321	681	831	1060
2	1006	790	536	1561
3	1767	369	118	1639

RESULTS AND DISCUSSION

CO-II Gene

The CO-II gene in all black flies had 229 codons, including an ATG initiation codon, and a complete putative TAA termination. The A + T bias at all codon positions (Table 2) was typical of that in other sequenced insects. Within black flies, there was amino acid variation in 27 codons (Fig. 1) of which 12 were phylogenetically informative. Outgroups varied at 73 amino acids.

From a functional standpoint, the amino acid substitutions observed in black flies are not surprising. Almost all amino acid substitutions noted in black flies also occur in other insects (Liu and Beckenbach, 1992). In some cases closely related species, with little nucleotide divergence, had amino acid replacements. Residue 42 in *S. confusum* is a conservative change (L for M). Residue 49 (T for K) in *S. jacumbae* might be considered a nonconservative replacement but would not greatly affect the hydropathic profile of this region. Such amino acid differences, even within a species, are not unusual (Willis *et al.*, 1992; Sperling and Hickey, 1994).

Weights

Although weighting of nucleotide characters is common, there seems little agreement as to how such weights should be derived and applied. Because we were interested in relationships at different taxonomic levels, we felt it desirable to retain third position nucleotides which, potentially, could have been the only positions differing between closely related species. It is also becoming evident that third positions, even though highly homoplastic, may contain most of the phylogenetic structure in a data set (e.g., Björlund, 1999). Although our weighting scheme resulted in relatively large weights for some types of changes, weights were similar to those derived by other methods in that the third position has lowest weight with transitions less than transversions. This pattern also holds at the first position, but C-T changes (primarily among leucine codons) receive a lower weight than any third position transversion. Relative weights assigned by position (ca. 4:14:1) are similar to those computed for a cytochrome *b* data set (4:17:1) by Arnason and Gullberg (1994) and further explored by Milinkovitch *et al.* (1996). Our approach, similar to that of Williams and Fitch (1990), further subdivided the position weights into specific nucleotide changes. This resulted in a departure from weights assigned by other workers for the second position, with transitions and transversions being similarly weighted. A second position C-G transversion in codon 174 was the only second codon position nucleotide which entered into parsimony analysis of the ingroup, the few other second position differences between black flies all being autapomorphies. Our weighting scheme often violated the triangle inequality. This would also be the case when applied to

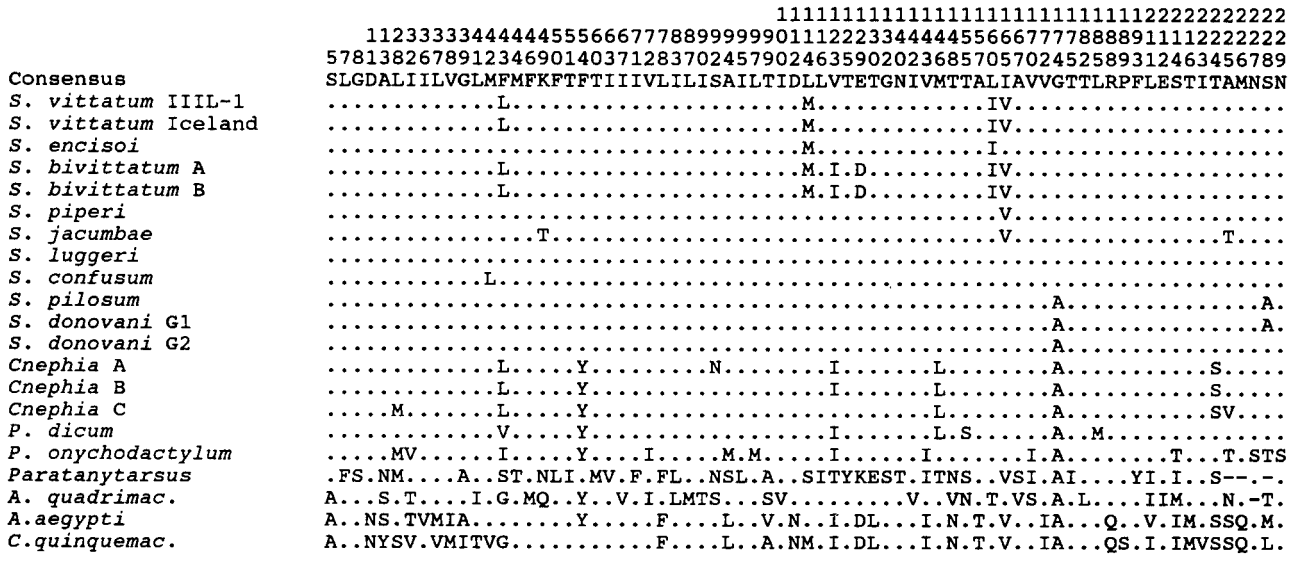

FIG. 1. Variable amino acids in CO-II in black flies and Nematocera outgroups.

TABLE 3

RASA (Relative Apparent Synapomorphy Analysis) Tests of Phylogenetic Signal

Taxon status	Observed slope	Null slope	tRASA	Significance
Unrooted, all taxa	5.466	5.535	−0.2318	NS
Rooted	8.183	5.003	7.7149	$P < 0.0005$
Rooted, pruned	2.371	2.794	−0.946	NS
Unrooted, outgroup taxa removed	3.974	3.455	1.768	$P < 0.025$
Rooted with *Paratanytarsus*, outgroup taxa removed	7.514	4.888	7.685	$P < 0.0005$
Unrooted, outgroup taxa removed, pruned	1.481	2.238	−5.338	$P < 0.0005$
Rooted with *Paratanytarsus*, pruned	2.153	2.647	−1.351	NS

Note. Values of $P < 0.05$ reject the hypothesis that the data set does not contain a significant phylogenetic signal.

other large data sets such as that of Liu and Beckenbach (1992) and we used weights as computed, but note that adjusting them to satisfy the triangle inequality did not alter tree topology. Cunningham (1997) found that six-parameter parsimony gave consistent positive results between congruence and accuracy.

Phylogeny

The data set contained significant phylogenetic signal. Even though the outgroup taxa were phylogenetically distant from the ingroup taxa, their contribution to unweighted signal content was appropriate to resolving relationships among the ingroup taxa (Lyons-Weiler *et al.*, 1998). However, uncontroversial relationships among closely related sister taxa was the only significant source of phylogenetic signal (Table 3). Each method of treebuilding produced a single, optimal tree (Figs. 2 and 3). Maximum likelihood and unweighted parsimony trees displayed disparate topologies (Fig. 3), but the maximum likelihood tree was congruent with the weighted maximum parsimony tree (Fig. 2) when the highly divergent outgroup taxa were removed from the maximum likelihood analysis. We cautiously propose Fig. 2 as the best estimate of phylogenetic relationships among the sampled taxa based on the CO-II gene, with reservations as to the correct ordering of nodes below those of closely related sister taxa. The best estimate of phylogenetic relationships implies a monophyletic *Simulium*, but the alternative hypotheses depict a paraphyletic *Simulium*.

Both Simuliidae and Culicidae were strongly supported as monophyletic families. CO-II sequences tend to support the finding of Pawlowski *et al.* (1996) that Chironomidae may be a sister group to other Nematocera. Pruess *et al.* (1992) found that mosquito, *Drosophila*, and black fly were about equally divergent in tRNA genes.

Nucleotide use at variable vs invariant sites differed significantly (χ^2, $P < 0.01$) at all codon positions. Although weights differed somewhat, both methods of weighting (nucleotide use at all vs only variable sites) resulted in identical single most parsimonious trees. The resultant parsimony phylogeny, with four non-black fly outgroup taxa (Fig. 2), fully agreed with current classification of black flies into species groups, but provided poorer resolution of relationships among species groups. Nodes at deeper levels were more weakly supported but resulted in phylogenies consistent with current classification at the subgeneric or lower levels. Molecular data suggest, as does morphology, the fossil record, and cytological evidence, an early divergence of currently recognized taxa at the subgeneric or deeper levels. Subsequent to this early radiation, there appears to have been little morphological differentiation.

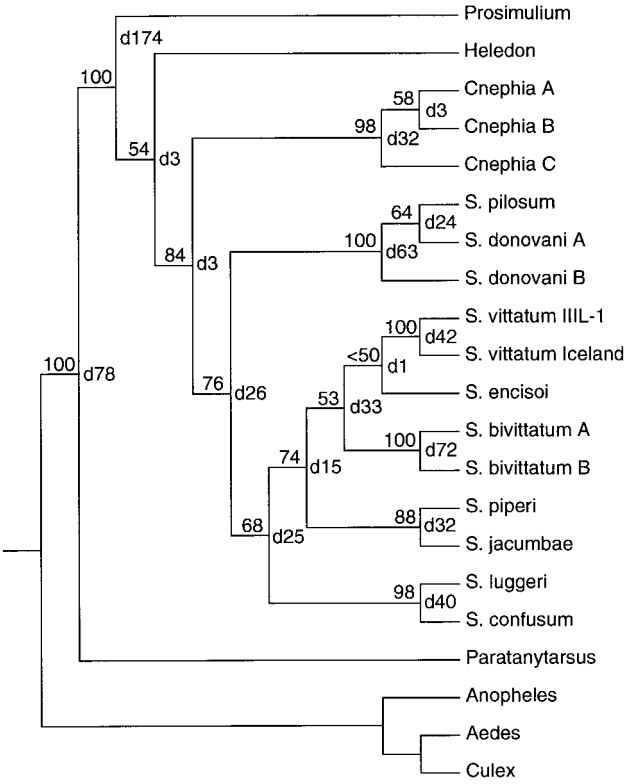

FIG. 2. Hypothesis of phylogenetic relationships by weighted parsimony (all taxa) and maximum likelihood (ingroup taxa and *Paratanytarsus* only). Percentage bootstrap support (500 replicates) is given to the left of each node; Bremer support values appear internal to their respective node and are preceded by the letter "d."

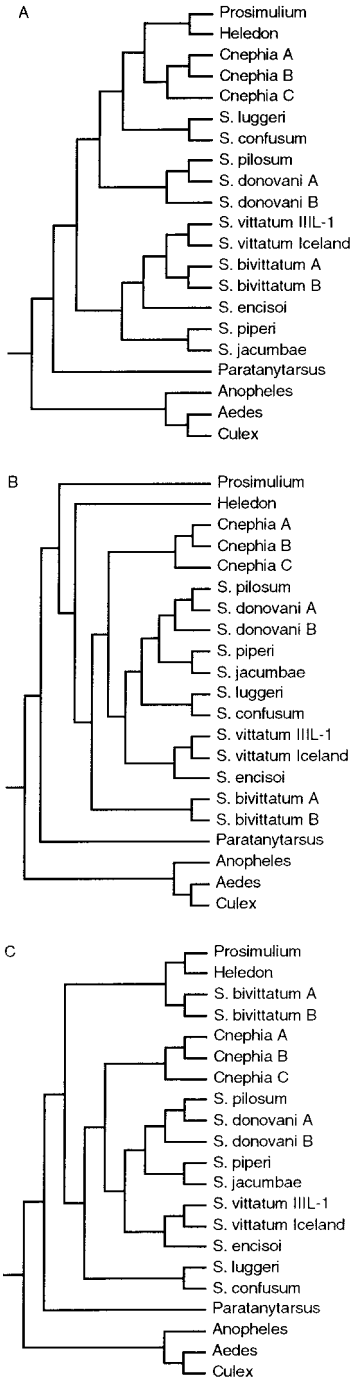

FIG. 3. Hypotheses of phylogenetic relationships by method of reconstruction: (A) maximum likelihood (all taxa); (B) neighbor-joining tree based on log-determinant distances; (C) unweighted parsimony, reweighted parsimony.

Prosimulium. Prosimuliini are generally accepted to be the more primitive black flies but there is no general agreement as to which genus, *Heledon* or *Prosimulium*, should be considered most plesiomorphic (Rothfels, 1979). The two Prosimuliini species examined here did not form a monophyletic clade. *P. onychodactylum* was as distant from *H. dicentum* as *H. dicentum* was from all other black flies. *P. onychodactylum* is a species complex (Henderson, 1986), and our sequence was probably derived from cytospecies 10.

Cnephia. Tribal placement of *Cnephia* is questionable. Crosskey and Howard (1997) place *Cnephia* in Prosimuliini. Sohn *et al.* (1975) and Teshima (1972), using DNA:DNA hybridization, found that *Cnephia* was as similar to some *Simulium* as species included in *Simulium* were to each other. Both Currie (1988), based on morphological analysis, and Moulton (2000), using molecular data, place *Cnephia* in Simuliini. Our analysis of the CO-II gene also provided support for inclusion of *Cnephia* basally in the Simuliini clade.

Our results pose an interesting taxonomic problem. *Cnephia* sp. A and sp. B, which differ by 7% unweighted sequence divergence, were initially assumed to be *C. pecuarum*, a serious livestock pest in Texas. *C. ornithophilia* in this area may occur sympatrically with *C. pecuarum* (Procunier, 1982), and thus *C. ornithophilia* may have been sampled. We were able to obtain a partial sequence from *C. ornithophilia* from South Carolina (where neither *C. dacotensis* nor *C. pecuarum* is known to occur). In 195 bp sequenced (codons 111–175), we found *C. ornithophilia* to differ by 4.6% unweighted sequence divergence from putative *C. dacotensis* from New Hampshire but by 7.2–12.3% from *C. pecuarum*. The situation was further complicated by two partial sequences we obtained from *C. dacotensis* from Nebraska where no other *Cnephia* are known. Both were different, with each being similar to one of the putative *C. pecuarum* sequences from Texas. At this time we cannot definitely ascribe the correct name to either sequence. The New Hampshire putative *C. dacotensis* did not match either of the two partial sequences obtained from Nebraska specimens. Thus, we have an anomalous situation in which we have four distinct haplotypes for three putative species, with two putative species each having two shared haplotypes. These results point out how little is known about levels of genetic divergence among sibling species, as well as the great range of genetic diversity represented by morphologically indistinguishable black flies.

Subgenus Eusimulium. *S. aureum* in North America is a complex of perhaps 10 sibling species (Dunbar, 1959; Leonhardt, 1985), none of which is conspecific with the European nominal species. Of the *Eusimulium* taxa we studied, we have confidence only in the name identification of *S. pilosum* (formerly *S. aureum* "B"). The two specimens labeled *S. donovani* A and *S. donovani* B came from a collection in which all individuals cytologically determined (Peter Adler, personal communication) were *S. donovani* (formerly *S. aureum* "G"). However, Adler (personal communication) found *S. pilosum* at this site in a previous year. The close

similarity to *S. pilosum* from Nebraska prompted us to sequence a second individual. Each differed from the other two by <1% sequence divergence. This could represent within-species variation or, at the other extreme, three different species. Leonhardt (1985) indicated that "B" and "G" were sister species in chromosomal phylogeny.

Subgenus Psilopelmia. The subgenus *Psilopelmia* is a complex of about 50 named species, primarily Neotropical, but with representatives in the Western Palearctic. *S. bivittatum* B came from a stream with fluctuating water levels and often carrying silt, in contrast to A which was from a clean sandy-bottom stream with stable flow, typical of the habitat in which *S. bivittatum* occurs in Nebraska.

Subgenus Psilozia. Based on cytological (Rothfels and Featherston, 1981) and ecological (Adler and Kim, 1984) evidence, *S. vittatum* is a complex of two sibling species, designated IS-7 and IIIL-1. *S. vittatum* from Iceland, lacking a diagnostic sex chromosome system, cannot be reliably assigned to either currently recognized sibling. Our sequences from IIIL-1 from Nebraska and a specimen from Iceland differed by less than 1%. Zhu *et al.* (1998) found that the two putative sibling species shared many restriction fragment length polymorphisms in the mitochondrial genome. Tang *et al.* (1996) failed to identify sibling-specific differences in either 12S or 16S mitochondrial ribosomal sequences. Zhu *et al.* (1998) identified sequence polymorphism in the ND4 gene, but the most common haplotype was present in both siblings. *S. encisoi* differed from *S. vittatum* by 9% with which it formed a poorly supported clade. *Psilopelmia* and *Psilozia,* both presumed to be of Neotropical origin, were supported as a clade. A single amino acid change (D in *Psilopelmia* at 129 vs E in other black flies) suggests an apomorphic character state in *Psilopelmia.* Had greater weight been given third position transversions which change amino acids, *S. encisoi* would more clearly have been placed in the *Psilozia* clade, a phylogenetic position supported by morphology.

Subgenus Simulium. The subgenus *Simulium* is divided into 18 species groups. *S. piperi* and *S. jacumbae,* members of the *hunteri* group, are morphologically distinct as larvae and pupae, but adults can be reliably separated only by genitalic characters. *S. piperi* has a more northern distribution. Where sympatric, both tend to occur in small, spring-fed streams which differ primarily in water temperature, *S. piperi* occurring in cooler streams. These two species, with *Psilozia* and *Psilopelmia,* form a clade, in agreement with the Nearctic or Neotropical distribution of all included species.

The two species placed in the *jenningsi* species group, *S. luggeri* and *S. confusum,* form a strongly supported clade, but its relationship to other species groups is unclear. In their revision of the *jenningsi* group, Moulton and Adler (1995) placed *S. luggeri* basally with *S. confusum* treated as derived. Included species in the *jenningsi* group are all eastern Nearctic in distribution. The sequence of a member of the eastern Palearctic or western Nearctic *S. malyschevi* species group, considered by Moulton and Adler to be a logical sister group, might help resolve the relationship of the *jenningsi* group to other black flies.

Our weighting method is readily computed from the data, providing an a priori estimate of weights, and requires little computer time, permitting bootstrap estimates of support. Although lower level relationships were weakly supported, those relationships were both logical and better supported than obtained by unweighted parsimony analyses and distance methods. Inclusion of more species, and better outgroup taxa, would be suggested to determine its full utility. The *Paratanytarsus* which we used is a highly derived parthenogenetic species. At the time this study was conducted, a specimen of Thaumaleidae was not available nor were we yet aware of the evidence for a sister group relationship with Simuliidae. Inclusion of additional taxa in Prosimuliini would perhaps aid in breaking up long branches (Graybeal, 1998) and better resolve the placement of *Cnephia.*

We applied this weighting method to other CO-II data sets. For insect orders (Liu and Beckenbach, 1992), we also found strong bootstrap support for monophyly of cockroach and termite (97%) and Hymenoptera (93%) but all other relationships remained poorly, and illogically, resolved. It is apparent that the CO-II gene is not informative at the insect order level, a condition that is unlikely to be solved by any method of weighting or analysis. The low differential in position weights (2:3:1) and transversion:transition weights (1.5:1, 1.5:1, 2:1) suggests that saturation was being approached at all positions.

Results for Collembola were almost identical to those reported by the authors (Frati *et al.,* 1997). Differential between position weights (3:5:1) and transversion: transition (1.5:1, 1:1, 3:1) suggested that although saturation was being approached, a strong phylogenetic signal was retained.

Our results for the beetle phylogeny of Emerson and Wallis (1995), with $tRNA_{LEU}$ excluded, supported their phylogeny for nodes with >60% bootstrap support in either tree, but arrangements differed considerably at less-supported nodes. We computed much greater transversion:transition weights than used by the authors at both first (18:1) and third (10:1) codon positions.

Our analysis of the mosquito data (Foley *et al.,* 1998) gave somewhat different results. Because the authors reported only relative improvement obtained by weighting, instead of bootstrap support, it is difficult to

TABLE 4

Frequency of Nucleotide Use at Variable Sites and Pairwise Changes by Codon Position in the Mitochondrial CO-II Gene for Black Flies and Outgroups

	Relative frequency of use at codon position		
Nucleotide	1	2	3
A	0.3363	0.2601	0.4536
C	0.1738	0.2010	0.0919
G	0.2143	0.1373	0.0311
T	0.2756	0.4017	0.4233
Change	Number of pairwise changes		
A ↔ C	209	208	1,181
A ↔ G	1044	102	1,651
A ↔ T	801	128	6,412
C ↔ G	11	266	97
C ↔ T	2046	366	5,290
G ↔ T	397	118	302
Total	4508	1188	14,933

make a direct comparison of results. We obtained the same sister taxa when bootstrap support by our analysis was greater than 60%, but deeper relationships were poorly resolved with branching quite different from that of the authors. Without knowing their actual bootstrap values, we cannot judge if the phylogenies differed significantly, nor can we compare the computed weights used by us with the unstated weights used by the authors. We computed 5:17:1 position weights and transversion:transition weights within position of 7:1, 1:1, and 4:1.

We conclude that the weighting method which we propose for this type of data will provide results as good as those obtained by any other method. This study supports results of other authors that the CO-II gene is likely to be useful for exploring relationships at subgeneric or lower levels, but is of questionable value at higher taxonomic levels.

Of more practical value is the contribution of molecular data as a source of characters for identification of the taxonomically challenging Simuliidae. The presence of sibling species is usually first suspected on the basis of chromosomal, or apparent ecological, differences. Cytology, unfortunately, permits identification only of mature larvae collected and preserved under ideal conditions and examined by persons with highly specialized skills. Molecular data, on the other hand, are applicable to all life stages and in black flies provide a comparatively rich source of variation relative to highly conserved morphology. Once discovered, molecular characters, ideally from independent loci, can then be used in combination with morphology and/or cytology to define species boundaries.

APPENDIX: COMPUTATION OF WEIGHTS

Weights were computed as expected/observed changes based on a stochastic model assuming base frequencies are at equilibrium at each codon position. Here we provide a simple algebraic solution illustrated with computation of codon position weights and weight assigned A ↔ C changes at codon position 1. Although each weight can be computed from a single equation, we will use a two-step process in which position weights are first computed, then multiplied by relative weights for each type of change within position. The example below is based on nucleotide frequencies at variable positions, but is equally applicable if one desires to include invariant sites.

The expected relative probability of change at each codon position, P_c, where f_{cn} is frequency of nucleotide n at codon position c, is

$$P_c = \frac{1 - \sum_{n=1}^{4} f_{cn}^2}{\sum_{c=1}^{3} (1 - \sum_{n=1}^{4} f_{cn}^2)}. \quad (1)$$

From Table 4, given 20,629 pairwise differences in the entire data set, expected frequency of changes at codon position is 0.7182/2.0309 = 0.3536. Observed frequency is 4508/20629 = 0.2185 and position weight, W_c, is expected/observed = 0.3536/0.2185 = 1.62. Note that if nucleotide frequency were the same at each position, then the expected changes at each position under a stochastic model would be 1/3.

TABLE 5

Weights Assigned to Nucleotide Changes Computed from Total Data Base Including Outgroups

	Arithmetic position weights			Integer position weights		
Change	1	2	3	1	2	3
A ↔ C	3.59	6.13	0.75	25	42	5
A ↔ G	1.23	4.71	0.15	8	32	1
A ↔ T	2.72	6.10	0.59	19	42	4
C ↔ G	50.08	5.30	0.65	343	36	4
C ↔ T	0.46	6.26	0.19	3	43	1
G ↔ T	4.03	7.32	0.89	28	50	6

Within codon position, expected relative frequency of each type of nucleotide change, $n_i \leftrightarrow n_j$, is given:

$$E(f_{c:i\leftrightarrow j}) = \frac{2f_{ci}f_{cj}}{1 - \sum_{n=1}^{4} f_{cn}^2} \qquad (2)$$

Again for codon position 1, observed frequency $f_{(c:i \leftrightarrow j)}$ of A ↔ C changes is 209/4508 = 0.0464, expected frequency is 2(0.2928)(0.1260)/0.7182 = 0.1027, and relative within-position weight $W_{(c:i \leftrightarrow j)}$ is 0.1027/0.0464 = 2.2142.

Weight for A ↔ C changes within codon position 1 is the product of the two weights:

$$W_{c(A \leftrightarrow C)} = W_c W_{(c:A \leftrightarrow C)} = 2.2142(1.62) = 3.5869.$$

For some phylogenetic analyses, only integer weights are permitted. These are readily derived from $W_{c(i \leftrightarrow j)}/W_{\min}$ as reported in Table 5.

Upon request, we can provide a program written in Microsoft GW-BASIC which permits rapid computation of weights.

ACKNOWLEDGMENTS

Peter Adler, Ronnie Byford, Gisli Gislason, Bai Xiong, and James Robinson provided some of the black flies used in this study. Roberta Weber provided the *Paratanytarsus*. We thank David Swofford for permission to use the test version of PAUP*. Linda Young and David Taylor provided constructive reviews and assistance in analysis. This study is published with approval of the Director as Journal Series 11874, Agricultural Research Division, University of Nebraska, and is supported in part by NIH Grant AI29456 to T.O.P.

REFERENCES

Adler, P. H. (1987). Ecology of black fly sibling species. *In* "Black Flies: Ecology, Population Management, and Annotated World List" (K. C. Kim and R. W. Merritt, Eds.), pp. 62–76. Pennsylvania State Univ., University Park, PA.

Adler, P. H., and Kim, K. C. (1984). Ecological characterization of two sibling species, IIIL-1 and IS-7, in the *Simulium vittatum* complex (Diptera: Simuliidae). *Can. J. Zool.* **62:** 1308–1315.

Arnason, U., and Gullberg, A. (1994). Relationship of baleen whales established by cytochrome *b* gene sequence comparison. *Nature* **367:** 726–728.

Beckenbach, A. T., Wei, Y. W., and Liu, H. (1993). Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. *Mol. Biol. Evol.* **10:** 619–634.

Björlund, M. (1999). Are third positions really that bad? A test using vertebrate cytochrome *b*. *Cladistics* **15:** 191–197.

Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795–803.

Bremer, K. (1994). Branch support and tree stability. *Cladistics* **10:** 295–304.

Brown, J. M., Pellmyr, O., Thompson, J. N., and Harrison, R. G. (1994). Phylogeny of *Greya* (Lepidoptera: Prodoxidae), based on nucleotide sequence variation in mitochondrial cytochrome oxidae I and II: Congruence with morphological data. *Mol. Biol. Evol.* **11:** 128–141.

Cooper, C. E., Nicholls, P., and Freedman, J. A. (1991). Cytochrome *c* oxidase: Structure, function, and membrane topology of the polypeptide subunits. *Biochem. Cell. Biol.* **69:** 586–607.

Crosskey, R. W. (1990). "The Natural History of Blackflies." Wiley, Chichester.

Crosskey, R. W., and Howard, T. M. (1997). "A New Taxonomic and Geographical Inventory of World Blackflies (Diptera: Simuliidae)." Natural History Museum, London.

Cunningham, C. W. (1997). Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Syst. Biol.* **46:** 464–478.

Currie, D. C. (1988). Phylogeny of primitive Simuliidae (Insecta: Diptera: Culicomorpha). Ph.D. dissertation, Univ. Alberta, Edmonton, Canada.

Dunbar, R. W. (1959). The salivary gland chromosomes of seven forms of black flies included in *Eusimulium aureum* Fries. *Can. J. Zool.* **37:** 495–525.

Emerson, B. C., and Wallis, G. P. (1995). Phylogenetic relationships of the *Prodontria* (Coleoptera: Scarabaeidae: subfamily Melolonthinae), derived from sequence variation in the mitochondrial cytochrome oxidase II gene. *Mol. Phylogenet. Evol.* **4:** 433–447.

Eriksson, T. (1997). AutoDecay ver. 2.9.7 (Hypercard Stack distributed by the author). Botaniska Institutionen, Stockholm Univ., Stockholm.

Foley, D. H., Bryan, J. H, Yeates, D, and Saul, A. (1998). Evolution and systematics of *Anopheles:* Insights from a molecular phylogeny of Australian mosquitoes. *Mol. Phylogenet. Evol.* **9:** 262–275.

Frati, F., Simon, C., Sullivan, J., and Swoffford, D. L. (1997). Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. *J. Mol. Evol.* **44:** 145–158.

Friedrich, M., and Tautz, D. (1997). Evolution and phylogeny of the Diptera: A molecular phylogenetic analysis using 28s rDNA sequences. *Syst. Biol.* **46:** 674–698.

Graybeal, A. (1998). Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47:** 9–17.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22:** 160–174.

Henderson, C. A. P. (1986). A cytological study of the *Prosimulium onychodactylum* complex (Diptera: Simuliidae). *Can. J. Zool.* **64:** 32–44.

Ho, C-M., Liu, Y-M., Wei, Y-H., and Hu, S-T. (1995). Gene for cytochrome *c* oxidase subunit II in the mitochondrial DNA of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). *J. Med. Entomol.* **32:** 174–180.

Kallersjo, M., Farris, J. S., Kluge, A. G., and Bult, C. (1992). Skewness and permutation. *Cladistics* **8:** 275–287.

Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. *J. Mol. Evol.* **29:** 170–179.

Leonhardt, K. G. (1985). A cytological study of species in the *Eusimulium aureum* group. *Can. J. Zool.* **63:** 2043–2061.

Liu, H., and Beckenbach, A. T. (1992). Evolution of the mitochondrial cytochrome oxidase II gene among ten orders of insects. *Mol. Phylogenet. Evol.* **1:** 41–52.

Lyons-Weiler, J. F. (1999). RASA 2.3 software and documentation for the Mac. http://test1.bio.psu.edu/LW/rasatext.html.

Lyons-Weiler, J., Hoelzer, G. A., and Tausch, R. J. (1996). Relative apparent synapomorphy analysis (RASA) I: The statistical measurement of phylogenetic signal. *Mol. Biol. Evol.* **13:** 749–757.

Lyons-Weiler, J., Hoelzer, G. A., and Tausch, R. J. (1998). Optimal outgroup analysis. *Biol. J. Linnean Soc.* **64:** 493–511.

Milinkovitch, M. C., Leduc, R. G., Adachi, J., Farnir, F., Georges, M., and Hasegawa, M. (1996). Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in Cetaceans. *Genetics* **144:** 1817–1833.

Mitchell, S. E., Cockburn, A. F., and Seawright, J. A. (1993). The mitochondrial genome of *Anopheles quadrimaculatus* species A: Complete nucleotide sequence and gene organization. *Genome* **36:** 1058–1073.

Moulton, J. K. (2000). Molecular sequence data resolves basal divergences within Simuliidae (Diptera). *Syst. Entomol.* **25:** 95–113.

Moulton, J. K., and Adler, P. H. (1995). Revision of the *Simulium jenningsi* species-group (Diptera: Simuliidae). *Trans. Am. Entomol. Soc.* **121:** 1–57.

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

Page, R. D. M. (1998). TreeView ver. 1.5 (distributed by the author). Division of Environmental and Evolutionary Biology IBLS, Univ. Glasgow.

Pawlowski, J., Szadziewski, R., Kmieciak, D., Fahrni, J., and Bittar, G. (1996). Phylogeny of the infraorder Culicomorpha (Diptera: Nematocera) based on 28S RNA gene sequences. *Syst. Entomol.* **21:** 167–178.

Procunier, W. S. (1982). A cytological study of species in *Cnephia* s. str. (Diptera: Simuliidae). *Can. J. Zool.* **60:** 2852–2865.

Pruess, K. P., Zhu, X., and 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). *J. Med. Entomol.* **29:** 644–651.

Rodriquez, F., Oliver, J. L., Martin, A., and Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142:** 485–501.

Rothfels, K. H. (1979). Cytotaxonomy of black flies (Simuliidae). *Annu. Rev. Entomol.* **24:** 507–539.

Rothfels, K., and Featherston, D. (1981). The population structure of *Simulium vittatum* (Zett.): The IIIL-I and IS-7 sibling species. *Can. J. Zool.* **59:** 1857–1883.

Saraste, M. (1990). Structural features of cytochrome oxidase. *Q. Rev. Biophys.* **23:** 331–366.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87:** 651–701.

Smith, J. J., and Bush, G. L. (1997). Phylogeny of the genus *Rhagoletis* (Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase II. *Mol. Phylogenet. Evol.* **7:** 33–43.

Sohn, U. I. K., Rothfels, K., and Straus, N. A. (1975). DNA–DNA hybridization studies in black flies. *J. Mol. Evol.* **5:** 75–85.

Sperling, F. A. H., and Hickey, D. A. (1994). Mitochondrial DNA sequence variation in the spruce budworm species complex (*Choristoneura:* Lepidoptera). *Mol. Biol. Evol.* **11:** 656–665.

Spicer, G. S. (1995). Phylogenetic utility of mitochondrial cytochrome oxidase gene: Molecular evolution of the *Drosophila buzzatii* species complex. *J. Mol. Evol.* **41:** 749–759.

Tang, J., Pruess, K., and Unnasch, T. R. (1996). Genotyping North American black flies by means of mitochondrial ribosomal RNA sequences. *Can. J. Zool.* **74:** 39–46.

Templeton, A. R. (1983). Convergent evolution and non-parametric inferences from restriction fragment and DNA sequence data. *In* "Statistical Analysis of DNA Sequence Data" (B. Weir, Ed.), pp. 151–179. Dekker, New York.

Teshima, I. (1972). DNA–DNA hybridization in blackflies (Diptera: Simuliidae). *Can. J. Zool.* **50:** 931–940.

Waddell, P. J., and Steel, M. A. (1997). General time-reversible distances with unequal rates across sites: Mixing Γ and inverse Gaussian distributions with invariant sites. *Mol. Phylogenet. Evol.* **8:** 398–414.

Wilkerson, R. C., Parsons, T. J., Albright, D. G., Klein, T. A., and Braun, M. J. (1993). Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: *Anopheles*). *Insect Mol. Biol.* **1:** 205–211.

Williams, P. L., and Fitch, W. M. (1990). Phylogeny determination using dynamically weighted parsimony method. *Methods Enzymol.* **183:** 615–626.

Willis, L. G., Winston, M. L., and Honda, B. M. (1992). Phylogenetic relationships in the honey bee (Genus *Apis*) as determined by the sequence of the cytochrome oxidase II region of mitochondrial DNA. *Mol. Phylogenet. Evol.* **1:** 169–178.

Wood, D. M., and Borkent, A. (1989). Phylogeny and classification of the Nematocera. *In* "Manual of Nearctic Diptera" (J. F. McAlpine, Ed.), Vol. 3, pp. 1333–1370. Canadian Government Publishing Centre, Hull, Quebec.

Zhu, X., Pruess, K. P., and Powers, T. O. (1998). Mitochondrial DNA polymorphism in a black fly, *Simulium vittatum* (Diptera: Simuliidae). *Can. J. Zool.* **76:** 440–447.