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.© 2000 Academic Press

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 superfamil

Chironoidea with Chironomidae (midge) 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, Prosimuliiini and Simuliiini, 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-
TABLE 1
Collection Data for Black Flies Sequenced

<table>
<thead>
<tr>
<th>Species</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hederon denticum Dyar &amp; Shannon</td>
<td>California, Napa Co., 22 March 1990</td>
</tr>
<tr>
<td>Prosimulium onychodactylum Dyar &amp; Shannon</td>
<td>Colorado, Jackson Co., 5 Aug. 1989</td>
</tr>
<tr>
<td>Simulium (Eusimulium) pilosum (Knowlton &amp; Rowe) (= S. aureum “B”)</td>
<td>Nebraska, Lancaster Co., 24 April 1989</td>
</tr>
<tr>
<td>S. (Eusimulium) donovani Vargas G1 (= S. aureum “G”)</td>
<td>California, Mono Co., 25 J une 1991</td>
</tr>
<tr>
<td>S. (Eusimulium) donovani G2</td>
<td>California, Mono Co., 25 J une 1991</td>
</tr>
<tr>
<td>S. (Psilozia) vittatum Zetterstedt sibling III-1</td>
<td>Nebraska, Lancaster Co., 25 March 1988</td>
</tr>
<tr>
<td>S. (Psilozia) vittatum “Iceland”</td>
<td>Iceland, R. Laxa, 31 Aug. 1990</td>
</tr>
<tr>
<td>S. (Psilozia) encisoi Vargas &amp; Diaz Najera</td>
<td>New Mexico, Dona Ana Co., 2 J une 1992</td>
</tr>
<tr>
<td>S. (Psilozia) vittatum</td>
<td>Nebraska, Loup Co., 12 Oct. 1990</td>
</tr>
<tr>
<td>S. (Psilozia) vittatum</td>
<td>Wyoming, Crook Co., 21 May 1994</td>
</tr>
<tr>
<td>S. (Psilozia) vittatum</td>
<td>Nebraska, Scottsbluff Co., 11 J uly 1989</td>
</tr>
<tr>
<td>S. (Psilozia) vittatum</td>
<td>Nebraska, Loup Co., 12 J uly 1990</td>
</tr>
<tr>
<td>S. (Psilozia) confusum Moulton &amp; Adler</td>
<td>Nebraska, Saline Co., 6 J uly 1990</td>
</tr>
</tbody>
</table>

Outgroups

| Paratanytarsus sp. (Chironomidae) | Pennsylvania, Chester Co., J 1 an. 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) |

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 $-70^\circ$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 C1-J-2797, 5'-CCTCGACGGTTATCCAGATTACC-3'; C2-J-3123, 5'-GGACTACAGATAGAGGCTCC-3'; C2-J-3407, 5'-CATCAATGATCTGAATGTTAT-3'; C2-J-3684, 5'-GTTCAATGTTGCGAAATTGTGGG-3'; C2-N-3380, 5'-TCAATATCATTGATGACCAAT-3'; C2-N-3686, 5'-CAGTTGGTTATAAAAATCTATGATTTCG-3'; A8-N-3926, 5'-AAATAAGGATCCTGGTATGAAATGAC-3'; and A6-N-4126, 5'-AGTCCTAAGAAATGTTCTTATCA-3'.

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

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 (Fratini 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.
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-32P 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 Smacut pBlueScript 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 35S-labeling.

To the extent possible, both strands were sequenced. Sequence of the opposite strand usually resolved problems due to compression in direct sequence. Multiple clones (3 or 4) were sequenced for each taxon to determine a consensus; a few instances of presumed Tag 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. M 76433 and AF 08357–AF 08373.

Phylogenetic Analysis

Published sequences for three mosquito species, Aedes aegypti (Ho et al., 1995), Anopheles quadrinmaculata (Mitchell et al., 1993), and Culex quinquefasciatus (Ho et al., 1995), were included as outgroup taxa. We also sequenced a chironomid, Paratanytarsus (Ho et al., 1995), were included as outgroup taxa. We also sequenced a chironomid, Paratanytarsus et al., (Ho et al., 1993), and Culex quinquefasciatus (Eriksson, 1997). Trees were visualized using TreeView version 1.5 (Page, 1998).

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; Källersjö et al., 1992) were calculated using AutoDecay version 2.9.7 (Eriksson, 1997). Trees were visualized using TreeView version 1.5 (Page, 1998).
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. jacobae 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; Sterling and Hick, 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., Bjöllund, 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 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

### TABLE 2

<table>
<thead>
<tr>
<th>Nucleotide Use by Codon Position in Black Flies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide</td>
</tr>
<tr>
<td>position 1</td>
</tr>
<tr>
<td>position 2</td>
</tr>
<tr>
<td>position 3</td>
</tr>
</tbody>
</table>

**Consensus**

- S. vittatum III-1
- S. vittatum Iceland
- S. encisoi
- S. bivittatum A
- S. bivittatum B
- S. piperi
- S. jacobae
- S. luggeri
- S. confusum
- S. pilosum
- S. donovi G1
- S. donovi G2
- Cnephas A
- Cnephas B
- Cnephas C
- P. dicum
- P. onychodactyllum
- Paracanthyurus
- A. quadriracemac
- A. aegypti
- C. quinquesacmac

**Weights**

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

**FIG. 1.** Variable amino acids in CO-II in black flies and Nematocera outgroups.
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 Simuliida, but the alternative hypotheses depict a paraphyletic Simuliida.

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 Nematocephleri. 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 ($\chi^2, P < 0.01$) 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.

![Diagram of phylogenetic relationships](Fig. 2)

**TABLE 3**

<table>
<thead>
<tr>
<th>Taxon status</th>
<th>Observed slope</th>
<th>Null slope</th>
<th>tRASA</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrooted, all taxa</td>
<td>5.466</td>
<td>5.535</td>
<td>-0.2318</td>
<td>NS</td>
</tr>
<tr>
<td>Rooted</td>
<td>8.183</td>
<td>5.003</td>
<td>7.7149</td>
<td>$P &lt; 0.0005$</td>
</tr>
<tr>
<td>Rooted, pruned</td>
<td>2.371</td>
<td>2.794</td>
<td>-0.946</td>
<td>NS</td>
</tr>
<tr>
<td>Unrooted, outgroup taxa removed</td>
<td>3.974</td>
<td>3.455</td>
<td>1.768</td>
<td>$P &lt; 0.025$</td>
</tr>
<tr>
<td>Rooted with Paratanytarsus, outgroup taxa removed</td>
<td>7.514</td>
<td>4.888</td>
<td>7.685</td>
<td>$P &lt; 0.0005$</td>
</tr>
<tr>
<td>Unrooted, outgroup taxa removed, pruned</td>
<td>1.481</td>
<td>2.238</td>
<td>-5.338</td>
<td>$P &lt; 0.0005$</td>
</tr>
<tr>
<td>Rooted with Paratanytarsus, pruned</td>
<td>2.153</td>
<td>2.647</td>
<td>-1.351</td>
<td>NS</td>
</tr>
</tbody>
</table>

Note. Values of $P < 0.05$ reject the hypothesis that the data set does not contain a significant phylogenetic signal.
**Prosimulium.** Prosimuliiini 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 Simulini. Our analysis of the CO-II gene also provided support for inclusion of Cnephia basally in the Simulini 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

---

**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.
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 Psiloelmia. The subgenus Psiloelmia 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 IIII-1. S. vittatum from Iceland, lacking a diagnostic sex chromosome system, cannot be reliably assigned to either currently recognized sibling. Our sequences from IIII-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. encisioides differed from S. vittatum by 9% with which it formed a poorly supported clade. Psiloelmia and Psilozia, both presumed to be of Neotropical origin, were supported as a clade. A single amino acid change (D in Psiloelmia at 129 vs E in other black flies) suggests an aposmorphic character state in Psiloelmia. Had greater weight been given third position transversions which change amino acids, S. encisioides 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. jacum, 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 Psiloelmia, form a clade, in agreement with the Neartic 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 basically 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. malychevi 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 Ceratophyllum.

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 Colembola 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 tRNALeu 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
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 $A \rightarrow C$ 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

$$P_c = \frac{1}{\sum_{n=1}^{4} f_{cn}^2}.$$

From Table 4, given 20,629 pairwise differences in the entire data set, expected frequency of changes at codon position is $0.3536$. Observed frequency is $0.2185$ and position weight, $W_c$, is $0.3536 / 0.2185 = 1.62$. 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 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

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Relative frequency of use at codon position</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>0.3363</td>
</tr>
<tr>
<td>C</td>
<td>0.1738</td>
</tr>
<tr>
<td>G</td>
<td>0.2143</td>
</tr>
<tr>
<td>T</td>
<td>0.2756</td>
</tr>
</tbody>
</table>

### TABLE 5

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

<table>
<thead>
<tr>
<th>Change</th>
<th>Arithmetic position weights</th>
<th>Integer position weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A ↔ C</td>
<td>3.59</td>
<td>6.13</td>
</tr>
<tr>
<td>A ↔ G</td>
<td>1.23</td>
<td>4.71</td>
</tr>
<tr>
<td>A ↔ T</td>
<td>2.72</td>
<td>6.10</td>
</tr>
<tr>
<td>C ↔ G</td>
<td>50.08</td>
<td>5.30</td>
</tr>
<tr>
<td>C ↔ T</td>
<td>0.46</td>
<td>6.26</td>
</tr>
<tr>
<td>G ↔ T</td>
<td>4.03</td>
<td>7.32</td>
</tr>
</tbody>
</table>
Within codon position, expected relative frequency of each type of nucleotide change, \( n_i \), is given:

\[
E \left( f_{ci \rightarrow cj} \right) = \frac{2f_{ci}f_{cj}}{1 - \sum_{n=1}^{4} f_{cn}^2}
\]

(2)

Again for codon position 1, observed frequency \( f_{ci \rightarrow cj} \) of A \( \rightarrow \) C changes is 209/4508 = 0.0464, expected frequency is 2(0.2928)(0.01260)/0.7182 = 0.1027, and relative within-position weight \( W_{c(i \rightarrow j)} \) is 0.1027/0.0464 = 2.2142.

Weight for A \( \rightarrow \) C changes within codon position 1 is the product of the two weights:

\[
W_{c(A \rightarrow C)} = W_{c}W_{(CA \rightarrow C)} = 2.2142(1.62) = 3.5869.
\]

For some phylogenetic analyses, only integer weights are permitted. These are readily derived from \( W_{c(i \rightarrow j)}/W_{c} \), 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 Gislaslon, Bai Xiong, and James Robinson provided some of the black flies used in this study. Roberta Weber provided the Paratanytarsus. We thank David Swoford 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


