J. N. Am. Benthol. Soc., 2009, 28(3):584–595

DOI: 10.1899/08-150.1
Published online 16 June 2009

Mitochondrial lineages and DNA barcoding of closely related species in the mayfly genus *Ephemerella* (Ephemeroptera:Ephemerellidae)

Laurie C. Alexander[1], Melanie Delion[2], David J. Hawthorne[3], AND William O. Lamp[4]

Department of Entomology, University of Maryland, College Park, Maryland 20742 USA

David H. Funk[5]

Stroud Water Research Center, Avondale, Pennsylvania 19311 USA

Abstract. We compared genetic lineages in the mayfly genus *Ephemerella* (Ephemeroptera:Ephemerellidae) identified from mitochondrial DNA (mtDNA) to current taxonomy in 9 morphological taxa, including 2 geographically widespread species, *Ephemerella invaria* (=*E. inconstans*, *E. rotunda*, *E. floripara*) and *Ephemerella dorothea* (=*E. infrequens*). Maximum likelihood and parsimony analyses of the mtDNA sequences placed *E. inconstans* and *E. invaria* in a well-supported clade; however, mean Kimura 2-parameter genetic distance between the lineages was high (5.2%) relative to distance within lineages (1.3%). The phylogenetic relationships of synonyms *E. rotunda* and *E. floripara* were not resolved, but estimates of mean genetic distance to *E. invaria* were high for both (8.5% and 11.6%, respectively). Populations of *E. dorothea* were grouped in 2 well-supported clades (12.9% mean divergence) that did not include the synonym *E. infrequens* (20.9% mean divergence, based on a single sample). A large genetic distance (18.6%) also was found between eastern and western populations of *Ephemerella excrucians*. Western samples of *Ephemerella aurivillii* were so genetically distant from all other lineages (32.2%) that doubt about its congeneric status is raised. mtDNA data have been useful for identifying genetic lineages in *Ephemerella*, but our results do not support use of cytochrome oxidase I (COI) as a DNA barcode to identify species in this genus because we also found evidence of incomplete mtDNA lineage sorting in this gene. Use of the barcoding gene rediscovered some old taxonomic problems in *Ephemerella*, a result that emphasizes the importance of completing empirical systematic description of species before using single-character systems for identification.

Key words: mayflies, aquatic insects, mtDNA, COI, DNA barcoding, integrative taxonomy.

Identifying mayfly species in the genus *Ephemerella* Walsh, 1862 (Ephemerellidae) is difficult even for taxonomic experts, who disagree about the interpretation of morphologic characters and the validity of some taxa. In a systematic revision of the genus, Jacobus and McCafferty (2003) formally confronted the issue of population variability by revising the North American species into ½ its original number. For example, 8 species with limited regional distributions were synonymized with the common widespread *Ephemerella invaria* Walker, 1853. However, ecologists and population geneticists working in the field with 3 *E. invaria* synonyms, *Ephemerella floripara* McCafferty, 1985, *Ephemerella rotunda* Morgan, 1911, and *Ephemerella inconstans* Traver, 1932 maintain that differences in emergence time, distribution, habitat, allozymes, and behavior support the original species status of these taxa (DHF and LCA, unpublished data; D. Lenat, retired, North Carolina Division of Water Quality, personal communication). We addressed this disagreement by sequencing a short region of the mitochondrial deoxyribonucleic acid (mtDNA) cyto-

[1] E-mail address: lalex@umd.edu

[2] Present address: Conservation GIS Lab, Conservation Ecology Center, Smithsonian's National Zoological Park, 1500 Remount Road, Front Royal, Virginia 22630 USA. E-mail: delionm@si.edu

[3] E-mail addresses: djh@umd.edu

[4] lamp@umd.edu

[5] dhfunk@stroudcenter.org

chrome oxidase gene subunit I (COI) from tissue samples taken from specimens examined in the 2003 morphological study by Jacobus and McCafferty. Fresh samples collected specifically for our study also were analyzed with mtDNA and allozyme markers. The final data set represented populations sampled from sites separated by distances up to 1700 km in eastern North America plus 3 western taxa and was used to estimate: 1) intraspecific genetic distances among species and synonyms within *Ephemerella* across its eastern range, and 2) the mitochondrial phylogeny of species closely related to *E. invaria*.

The COI gene was selected because mtDNA is characterized by a fast mutation rate, limited recombination, and (usually) maternal inheritance, factors that contribute to the rapid accumulation of genetic polymorphisms in populations and species (Avise 2004). The major drawback to using mtDNA for broad inferences about species identity or diversity is that single-gene lineages (gene genealogies) capture only one small part of a species' genetic history and can overestimate or underestimate species diversity in unpredictable ways (Funk and Omland 2003, Rubinoff 2006). Furthermore, the unique genomic characteristics of mtDNA might limit, rather than enhance, its utility as a genetic marker in population and species research because the idiosyncratic biology of the mitochondrial organelle creates greater potential for discord between the evolutionary histories of the mtDNA genome and the species as a whole (Shaw 2002, Ballard and Whitlock 2004). Despite these shortcomings, mtDNA can be an informative source of characters for estimating phylogenies of recently diverged taxa (Moore 1995).

We also evaluated the utility of DNA barcoding for identifying morphologically ambiguous *Ephemerella* larvae to species without direct use of taxonomic keys or expertise (Hebert et al. 2003a). The term *barcoding gene* loosely describes a fragment of DNA that has low sequence divergence within species but high divergence among species, from which unknown samples can be placed accurately into species groups simply by calculating their pairwise genetic distances (Hebert et al. 2003a, Ball et al. 2005; but see also Will and Rubinoff 2004). The Consortium for the Barcode of Life (CBOL; http://barcoding.si.edu/) has recommended the mitochondrial COI gene as its standard diagnostic sequence for DNA barcoding of animal species globally. Short polymorphic regions flanked by highly conserved DNA priming sites make this gene easy to sequence in a wide range of taxa (Simon et al. 1994) and, thus, attractive as a standard locus for large-scale DNA barcoding.

DNA barcoding has shown promise as a tool for solving some serious taxonomic problems in aquatic entomology and stream bioassessment. It provides tools and data for automated, online identification of described taxa, even from immature or damaged specimens. Major efforts are underway to construct reference libraries of specimen data across all taxa (Ratnasingham and Hebert 2007, iBOL 2003), and a large number of samples from aquatic insect taxa already have been barcoded and catalogued (CBOL 2003, Trichoptera BOL 2007). In addition to facilitating identification, easily accessed databases of standardized samples from a large number of geographic locations document the distribution of genetic diversity at the sequenced locus and provide spatial data on a scale not previously available. Recent work using mtDNA sequence data in stream community ecology (Pfenninger et al. 2007), population genetics (Finn et al. 2007, Lytle et al. 2008), and systematics and taxonomy (Rach et al. 2008) illustrate the potential for new research opportunities that such databases might provide. Smaller scale, but highly practical, applications of DNA barcoding use standardized DNA to identify fish products (Smith et al. 2008), agricultural pests (Armstrong and Ball 2005), and disease vectors (Nelson et al. 2007). Here we apply barcoding methods to the mtDNA sequences generated for the phylogenetic study to evaluate its utility in our test assemblage of *Ephemerella*.

Study system

The family Ephemerellidae consists of 2 subfamilies, Ephemerellinae and Timpanoginae (McCafferty and Wang 2000). The subfamily Ephemerellinae and genus *Ephemerella*, the largest genus in Ephemerellidae, have undergone frequent revision in recent decades in North America (Allen and Edmunds 1962, 1963, 1965, Allen 1980, 1984, Jacobus and McCafferty 2003). *Ephemerella*, in particular, has been notable for problems of paraphyly, poor diagnostics, and high population-level variability in some species. These problems prompted the recent revisionary contribution by Jacobus and McCafferty (2003). Ephemerellids have distributions and ecologies that are favorable to current research topics in stream ecology, including toxicology, nutrient transport and cycling, insect dispersal, stream recolonization, and predator–prey interactions (McShaffrey and McCafferty 1991, Benke and Jacobi 1994, Rezanka and Hershey 2003, Beketov 2004). In addition, they are used routinely in freshwater biomonitoring efforts and are among some of the most sensitive aquatic insects in North America (Hilsenhoff 1987, Lenat 1993). Therefore, accurate definition of taxonomic and geographic species boundaries is important to fields of research outside of systematics

because cryptic genetic diversity represents a source of uncontrolled experimental error and a source of vital information for understanding evolutionary and ecological processes.

Methods

Taxon sampling

Seventy-eight unique mtDNA haplotypes representing 12 species and synonyms from 13 geographic regions were obtained (Table 1). This count does not include specimens collected in the same region that had nucleotide sequences diverged by <0.4%. Sequences with <380 base pairs (bp) of high-quality nucleotide sequence also were dropped, even though this step reduced 2 taxa (*Ephemerella infrequens* McDonnough, 1924, *E. floripara*) to a single sample each.

Preserved tissue samples from specimens used in the systematic revision of the genus (Jacobus and McCafferty 2003) were obtained from L. M. Jacobus (Indiana University). Additional tissue samples and voucher specimens were obtained from collections held at the Stroud Water Research Center (Avondale, Pennsylvania) and at the North Carolina Division of Water Quality (Raleigh, North Carolina). Four additional sequences were obtained from GenBank (Table 1).

Fresh samples of *Ephemerella* from Maryland, Virginia, Pennsylvania, and North Carolina were collected and preserved in 100% ethanol, stored at ambient temperature during transit, and put into long-term storage at −20°C. Heads were removed for DNA extraction, and thoraces and abdomens were labeled and stored as vouchers at the University of Maryland, Department of Entomology, and at Stroud Water Research Center.

Three sequences from other ephemerellid genera (*Attenella*, *Dannella*, and *Eurylophella*; Table 1) were downloaded from the Barcode of Life Data system (BOLD; Ratnasingham and Hebert 2007) for rooting the trees and estimating intergeneric genetic distances.

Specimen processing

DNA was extracted using the DNEasy Kit and protocol (Qiagen, Chatsworth, California). A small (380–490 bp) segment of the mtDNA COI gene that overlaps the region recommended for DNA barcoding was amplified by polymerase chain reaction (PCR) with primers 'Ron' (C1-J-1751, 5′-GGA TCA CCT GAT ATA GCA TTC CC-3′, 23 bp) and 'Nancy' (C1-N-2191, 5′-CCC GGT AAA ATT AAA ATA TAA ACT TC-3′, 26 bp) (Simon et al. 1994).

The cycling profile began with 1 cycle of DNA denaturation at 94°C for 2 min, followed by 35 cycles of sequence amplification (DNA denaturation at 94°C for 30 s, primer annealing at 47°C for 30 s, and sequence extension at 72°C for 1 min). PCR products were treated with exonuclease I and shrimp alkaline phosphatase to degrade unincorporated primers and deoxyribonucleotide triphosphates (dNTPs). The sequencing reactions were carried out using ABI BigDye® v3.1 terminators and the resulting products were sequenced on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, California). Sequencing reaction mixes contained 25 mol template, 1.25 pmol labeled primer, 2.75 mM $MgCl_2$, 10 mM Tris-HCl, pH 9.2, 100 mM KCl, 0.01 U pyrophosphatase, 1.4 μg *Taq* polymerase, 125 μM each dNTP and either ddATP, ddGTP, ddCTP, or ddTTP at 1 μM in a total volume of 20 μL. Thermal cycling consisted of 25 cycles of 10 s at 96°C, a 1°C/s ramp to 50°C, 15 s at 50°C, a 1°C/s ramp to 60°C, and 4 min at 60°C. Chromatographs of each sequence were examined to determine sequence quality and aligned using Sequencher (Gene Codes Inc., Ann Arbor, Michigan). Sequences obtained from forward and reverse primers were compared when needed to check base calls and confirm positions of polymorphic sites. The sequences were then edited in BioEdit (Hall 1999) to create nucleotide data matrices.

Phylogenetic analysis

Attenella attenuata (McDunnough), 1925 was selected as the outgroup for this analysis based on arguments in Jacobus and McCafferty's (2008) revision of Ephemerellidae genera.

Maximum parsimony and maximum likelihood analyses of the nucleotide matrix were conducted in PAUP* version 4.0b10 (Swofford 1998). Unweighted maximum parsimony analysis was done with heuristic searches using the tree bisection and reconnection (TBR) method of branch swapping (100 sequence-addition replicates). To assess the level of branch support, 1000 bootstrap replications were done with the same search and optimization criteria, except that the number of random sequence-addition replicates was reduced to 25. The maximum likelihood analysis used equally weighted trees from the parsimony analysis as starting points to estimate the log likelihood of trees obtained under a Hasegawa–Kishino–Yano + invariant + gamma (HKY + I + G) model of evolution (Hasegawa et al. 1985), with among-site rate variation modeled as a γ distribution with 4 rate categories. The best-fit model (HKY + I + G; $\alpha = 0.8965$, $I = 0.5133$) was selected through a hierarchical likelihood ratio test on the Modeltest 3.07

Table 1. Source of *Ephemerella* spp. specimens or data for analyses. BOLD = Barcode of Life Data system, DE = Delaware, FL = Florida, ID = Idaho, MD = Maryland, ME = Maine, MT = Montana, NC = North Carolina, NE = Nebraska, ON = Ontario, PA = Pennsylvania, TN = Tennessee, VA = Virginia, WV = West Virginia, Co = county, Jeff.NF = Jefferson National Forest. An asterisk (*) denotes the outgroup taxon.

Species	Locality	Date	Collector	GenBank accession number	BOLD identifier	Tree label
Specimens						
E. aurivillii	MT: Sweet Grass Co, Sweet Grass Creek	10 June 2000	LM Jacobus			*E. aurivillii* MT
E. catawba	NC: McDowell Co, Reedy Branch	21 April 2005	W Crouch			*E. catawba* NC1
E. catawba	NC: Haywood Co, Big Creek	12 June 2003	LM Jacobus			*E. catawba* NC2
E. dorothea dorothea	TN: Blount Co, pond in Cades Cove	13–20 May 2001	LM Jacobus			*E. dorothea* TN
E. dorothea	PA: Chester Co, White Clay Creek	1 May 2005	D Funk			*E. dorothea* PA
E. dorothea	MD: Appomattox Co, Fishpond Creek	March 2001	L Alexander			*E. dorothea* VA
E. dorothea	MD: Howard Co, South Stream	April 2003	L Alexander			*E. dorothea* MD2
E. dorothea	MD: Carroll Co, Morgan Run	20 March 2005	L Alexander			*E. dorothea* MD1
E. dorothea	NC: McDowell Co, Roses Creek	20 April 2005	W Crouch			*E. dorothea* NC
E. excrucians	NE: Brown Co, Long Pine Creek	6 June 2000	LM Jacobus			*E. excrucians* NE
E. excrucians	FL: Okaloosa Co, Turkey Creek	12 April 2001	LM Jacobus			*E. excrucians* FL
E. floripara	NC: Caldwell Co, Wilson Crab Gorge	April 2003	D Lenat			*E. floripara* NC
E. hispida	TN: Sevier Co, Dunn Creek	17 May 2001	LM Jacobus			*E. hispida* TN
E. inconstans	DE: Pratts Branch	11 April 1988	D Funk			*E. inconstans* DE
E. inconstans	TN: Anderson Co, Clinch River tributary	21 May 2001	LM Jacobus			*E. inconstans* TN
E. inconstans	MD and VA: multiple sites	2001–2004	L Alexander			*E. inconstans* MD VA
E. infrequens	ID: Valley Co, East Fork Salmon River	8–14 July 1989	D Funk			*E. infrequens* ID
E. invaria	NY: Delaware Co, W. Delaware River	13 May 2005	D Funk			*E. invaria* NY
E. invaria	PA: Berks Co, Angelica Creek	6 April 2005	D Funk			*E. invaria* PA1
E. invaria	PA: Chester Co, White Clay Creek	5 April 2005	L Alexander			*E. invaria* PA2
E. invaria	VA: Appomattox Co, Saunders Creek	12 April 2002	L Alexander			*E. invaria* VA
E. invaria	MD: Carroll Co, Joe Branch	20 March 2005	L Alexander			*E. invaria* MD
E. invaria	NC: Caldwell Co, Wilson Crab Gorge	April 2003	D Lenat			*E. invaria* NC2
E. invaria	NC: Caldwell Co, Wilson Creek	14 April 2005	D Lenat			*E. invaria* NC1
E. rotunda	PA: Berks Co, Manatawny Creek	6 April 2005	D Funk			*E. rotunda* PA
E. rotunda	NY: Delaware Co, W. Delaware River	13 May 2005	D Funk			*E. rotunda* NY

TABLE 1. Continued.

Species	Locality	Date	Collector	GenBank accession number	BOLD identifier	Tree label
E. rotunda	VA: Giles Co, North of Pembroke, Jeff.NF	11 March 2002	LM Jacobus			*E. rotunda* VA
E. rotunda	MD: Frederick Co, Fishing Creek	19 March 2001	L Alexander			*E. rotunda* MD
E. rossi	NC: Transylvania Co, Big Bearpen Branch	22 April 2005	W Crouch			*E. rossi* NC1
E. rossi	NC: Transylvania Co, Bearwallow Creek	21 April 2005	W Crouch			*E. rossi* NC2
E. subvaria	PA: Chester Co, White Clay Creek	5 April 2005	L Alexander			*E. subvaria* PA
NC species a	NC: Richmond Co, Naked Creek	April 2005	D Lenat			NC sp. a
Data						
E. dorothea				AY326813		*E. dorothea* WV
E. invaria				AY326814		*E. invaria* ME
E. subvaria				AY326815		*E. subvaria* ON
E. subvaria				AY326914		*E. subvaria* ME
*Attenella attenuata**					SBEP115-03|752_DR_PA	*A. attenuata* PA
Dannella simplex					SBEP114-03|749_WCC_PA	*D. simplex* PA
Eurylophella temporalis					SBEP131-03|803_ML_ME	*E. temporalis* ME

software (Posada and Crandall 1998). As in the parsimony analysis, the TBR method of branch swapping was used. Maximum likelihood bootstrap analysis (1000 replicates) was conducted with GARLI (version 0.951; Zwickl 2006) with the model parameters from Modeltest.

Barcoding analysis

Pairwise comparisons of the sequences and genetic distances within and among populations were estimated using the Kimura 2-parameter (K2P) method in the software program DNADIST (PHYLIP version 3.5c; J. Felsenstein, University of Washington, Seattle, Washington; http://evolution.genetics.washington.edu/phylip.html). Sequences of *Dannella simplex* (McDunnough), 1925 and *Eurylophella temporalis* (McDunnough), 1924 from the BOLD database (Ratnasingham and Hebert 2007) were included to estimate congeneric distances. A neighbor-joining (NJ) 50% majority-rule consensus tree was constructed in PAUP* version 4.0b10 (Swofford 1998) over 1000 bootstrap replicates. Taxa for which only a single haplotype was available were dropped from the NJ analysis, except for *A. attenuata*, which was used to root the NJ majority-rule tree.

Results

Maximum parsimony analysis

Within samples of *Ephemerella* (excluding other genera), 144 of the sequenced bps were parsimony-informative characters. Of these, 129 (89.5%) occurred in the 3rd codon position, 15 in the 1st codon position, and 1 in the 2nd codon position. Overall base frequencies were slightly biased towards adenine and thiamine (A + T; 56.5%), which is typical for insect mitochondrial genomes (Simon et al. 1994). A χ^2 test showed that bp frequencies were homogeneous across all taxa ($p = 1.0$). In the parsimony analysis, 17 equally parsimonious trees of length 779 were obtained. Strict consensus of the 17 most parsimonious trees (Fig. 1) showed strong bootstrap support (92%) for grouping *E. inconstans* with *E. invaria*, but a 2nd clade of haplotypes morphologically identified as *E. invaria* that was nested within the 1st clade also had strong bootstrap support (84%). The relationships of 2 other *E. invaria* synonyms, *E. rotunda* and *E. floripara*, were not well resolved by our analysis. Based on a single sample, *E. floripara* was placed with *Ephemerella dorothea* Needham, 1908, but without bootstrap support. Samples of *E. rotunda* from Pennsylvania

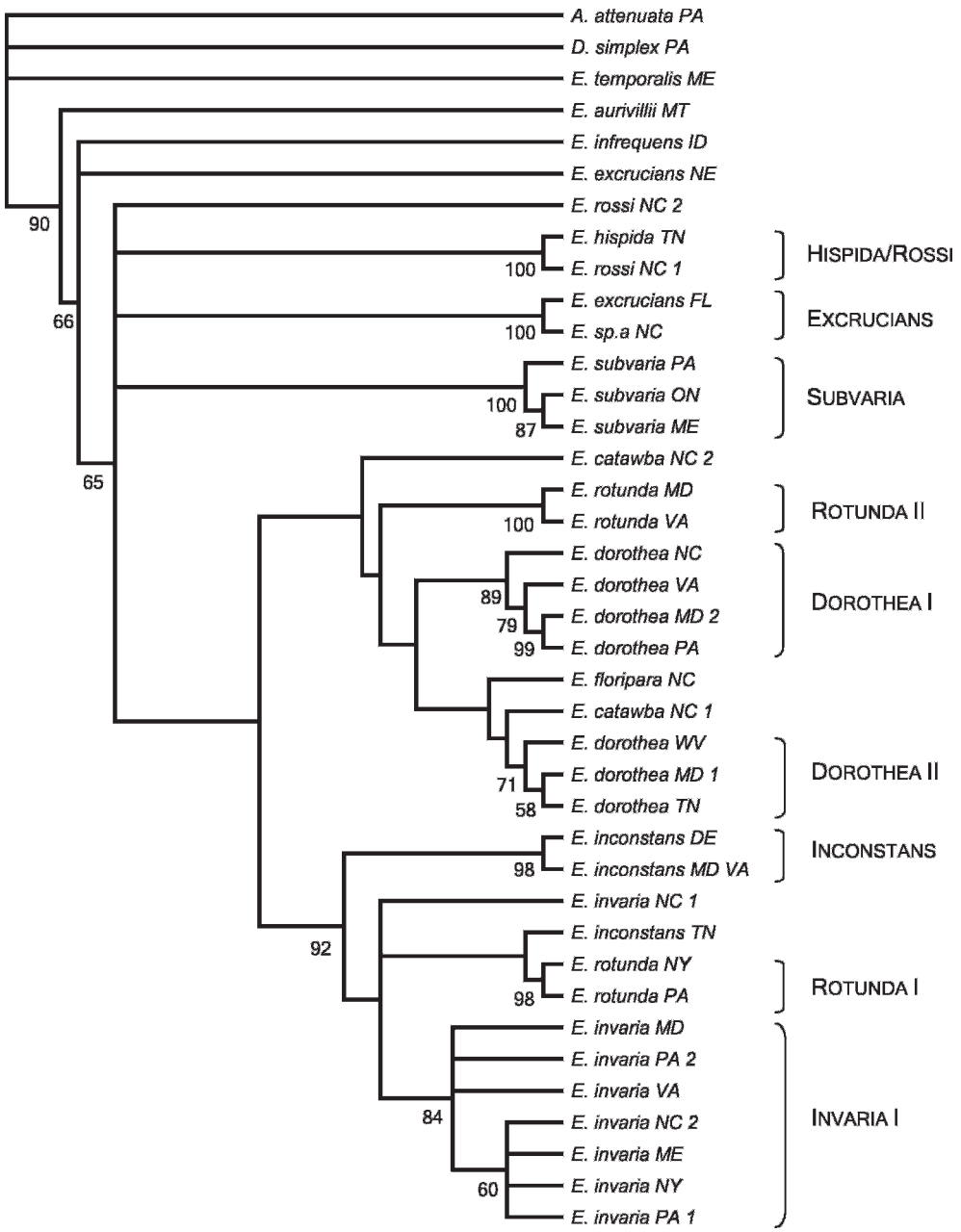

FIG. 1. Strict consensus of 17 equally parsimonious trees. Values shown are bootstrap support (1000 bootstrap replicates, values <50% not shown). Clade names (SMALL CAPS) denote lineages in Tables 2 and 3. *E.* = *Ephemerella* except *E. temporalis* = *Eurylophella temporalis*; *A. attenuata* = *Attenella attenuata*, *D. simplex* = *Dannella simplex*.

and New York grouped with *E. invaria*, but other samples from Maryland and Virginia did not. Specimens identified as *E. dorothea* were grouped into 2 clades (89% and 71% bootstrap support, respectively) with overlapping geographic ranges. The synonymy of *E. infrequens* with *E. dorothea* (Jacobus and McCafferty 2003) was not supported by our analysis. Although based on a single sample, the sequence divergence from the 2 *E. dorothea* clades (20.9% mean K2P genetic distance; Table 2) was large enough to call into question its current status as a synonym.

Likelihood analysis

The hypothesized relationships from the maximum likelihood analysis (Fig. 2) were identical to the maximum parsimony results (Fig. 1). The level of bootstrap support was similar in all branches (>60% support).

TABLE 2. Kimura 2-parameter genetic distances within and among all taxa. Values on the diagonal are average distances within a lineage; distances below diagonal are pairwise distances among lineages. Mean within-lineage distance is 2.9 ± 0.7%. Mean distance among taxa is 15.4 ± 1%. See Table 3 for data related to INVARIA II.

Number	Branch name	1	2	3	4	5	6	7	8	9	10	11	12	13
1	INVARIA II	(Table 3)												
2	DOROTHEA I	13.5	6.4											
3	DOROTHEA II	15.9	12.9	7.0										
4	ROTUNDA II	12.9	11.8	14.1	2.6									
5	HISPIDA/ROSSI	13.0	12.2	14.6	12.2	0.5								
6	*E. rossi* NC 2	16.9	17.6	17.7	19.3	15.9	–							
7	SUBVARIA	14.0	15.7	14.8	16.0	14.6	19.3	2.4						
8	*E. catawba* NC2	11.5	14.9	15.7	13.0	13.4	20.1	15.1	–					
9	*E. floripara* NC	11.6	13.3	12.9	11.8	12.4	19.3	16.6	14.4	–				
10	EXCRUCIANS	15.9	19.2	18.2	17.3	15.4	18.1	17.1	15.4	15.5	3.8			
11	*E. excrucians* NE	18.9	20.9	17.6	16.9	20.5	19.8	20.4	15.4	17.0	18.6	–		
12	*E. infrequens* ID	18.7	21.3	20.4	20.1	19.4	21.2	21.7	17.6	20.6	18.5	15.3	–	
13	*E. aurivillii* MT	32.7	34.1	34.3	33.3	32.4	33.9	37.0	29.7	32.7	31.1	29.5	25.1	–

DNA barcoding and distance analysis

The A + T content of publicly accessible sequences of *Ephemerella* species in the BOLD database (average COI sequence length = 629 bp, $n = 8$) was similar to the composition of sequences evaluated for our study (55.4% vs 56.5%).

The NJ 50% majority-rule consensus tree is shown in Fig. 3. Except for one haplotype (*E. dorothea* NC), groupings are the same as in the other trees (Figs 1, 2) in branches with >50% bootstrap support.

Mean within-lineage K2P genetic distance was 2.9 ± 0.7%; mean among-lineage genetic distance was 15.4 ± 1.1%, where *lineage* is a clade or unresolved branch of the maximum parsimony strict consensus tree in Fig. 1. In our analysis, bootstrap support was strong (90%) for the monophyly of *E. invaria* and *E. inconstans*. However, average K2P genetic distance among lineages in this clade was 5.2 ± 0.3%, which is high relative to the average within-lineage genetic distance of 1.3 ± 0.4% (Table 3). The nested clade labeled "INVARIA I," supported with a bootstrap value of 78%, had a geographic range extending ~1400 km from north to south but average genetic distance of just 1.3% (Table 3) among sample sites (Maine, New York, Pennsylvania, Maryland, Virginia, and North Carolina). By contrast, samples of "INVARIA I" and samples of the lineage labeled "INCONSTANS" that were collected from the same stream reaches in Maryland and in Virginia had an average genetic distance of 5.1% (Table 3), indicating that the differences observed here were not just the result of geographic distance. The specimens identified morphologically as *E. rotunda*, a new synonym of *E. invaria* (Jacobus and McCafferty 2003), consists of 2 genetic lineages labeled "ROTUNDA I" and "ROTUNDA II" that have diverged significantly (mean genetic distance = 12.6%) and are not monophyletic with respect to other recognized species, including *E. dorothea* and *Ephemerella subvaria* McDonnough, 1931. In addition, large sequence divergences were found between 2 lineages of *E. dorothea* (12.9% mean K2P distance, Table 2); between eastern and western populations of *Ephemerella excrucians* Walsh, 1862 (18.6% mean K2P distance); and from all lineages to the western haplotype of *Ephemerella aurivillii* (Bengtson), 1908 (32.2% mean K2P distance). A plot of the frequency distribution of pairwise genetic distances between all individuals (Fig. 4A) showed considerable overlap of intra- and interspecific variation. Distances among some taxonomic species fall within the distribution of congenerics, and distances of all taxa to the species *E. aurivillii* fall within the distribution of confamilials (Fig. 4A).

Discussion

We constructed phylogenetic trees of mtDNA COI lineages from geographically widespread populations of *Ephemerella* for comparison to morphologic species. We also estimated genetic distances among the mtDNA lineages to evaluate the use of distance-based COI barcoding to identify a limited number of morphologically ambiguous species in the genus. We found that the mtDNA gene trees are roughly consistent with described species. However, the presence of well-supported clades diverged by 4 to 16% in *E. invaria* and *E. dorothea* (Tables 2, 3) suggests

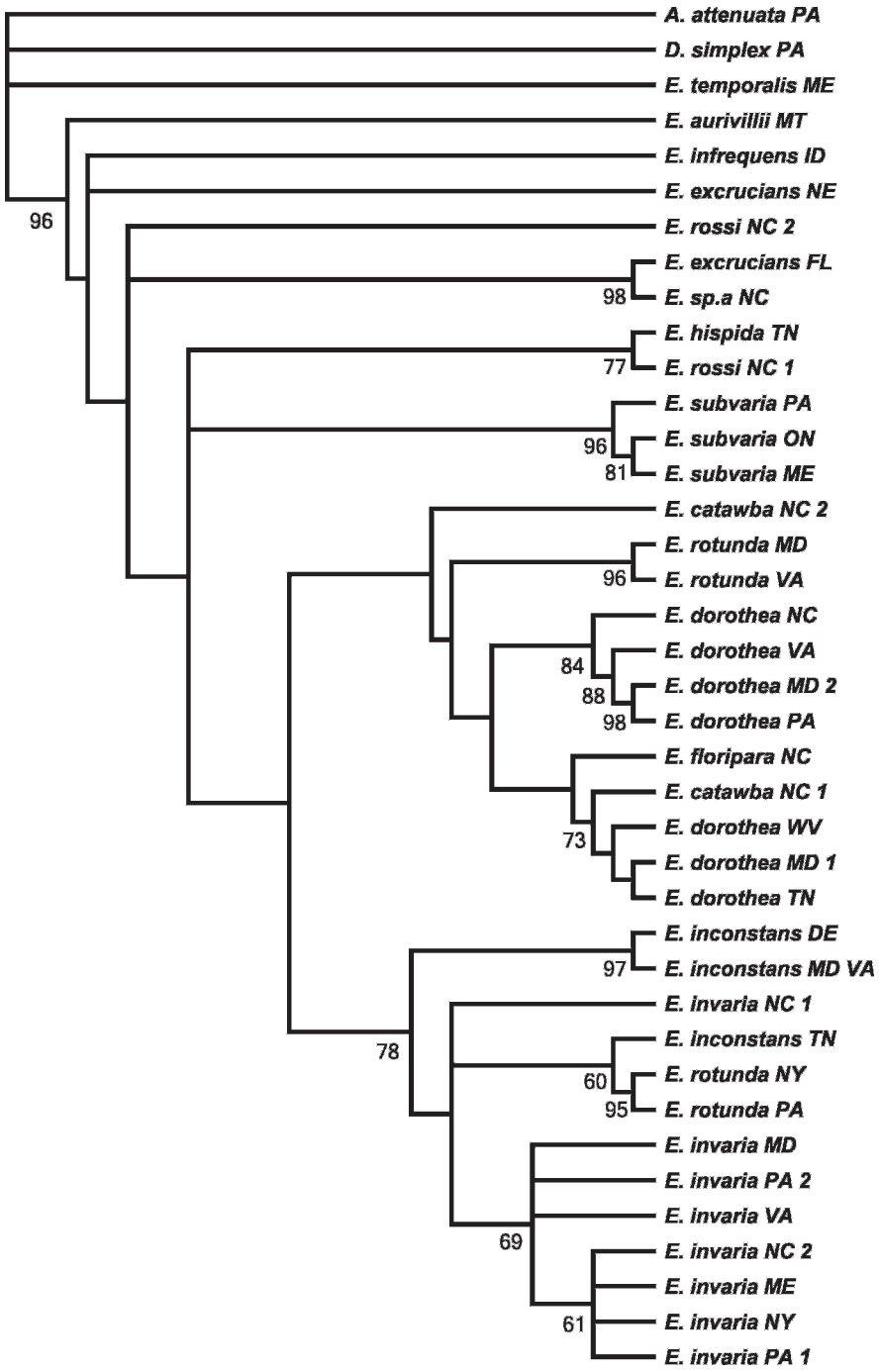

Fig. 2. Strict consensus maximum likelihood tree based on Hasegawa–Kishino–Yano + invariant + gamma (HKY + I + G) model of evolution. Values shown are bootstrap support (1000 bootstrap replicates, values <50% not shown). *E.* = *Ephemerella* except *E. temporalis* = *Eurylophella temporalis*; *A. attenuata* = *Attenella attenuata*, *D. simplex* = *Dannella simplex*.

that current taxonomy underestimates the diversity of these common mayfly taxa. Furthermore, lack of support for the synonymy of *E. floripara* and *E. rotunda* with *E. invaria*, deep divergence of eastern *E. dorothea* from its western synonym *E. infrequens*, and paraphyly of 3 well-described species (*Ephemerella rossi* Allen and Edmunds, 1965, *Ephemerella hispida* Allen and Edmunds, 1965, and *Ephemerella catawba* Traver, 1932) make a case for a systematic review of the genus that should include characters from nuclear genes and ecology in addition to morphology and mtDNA.

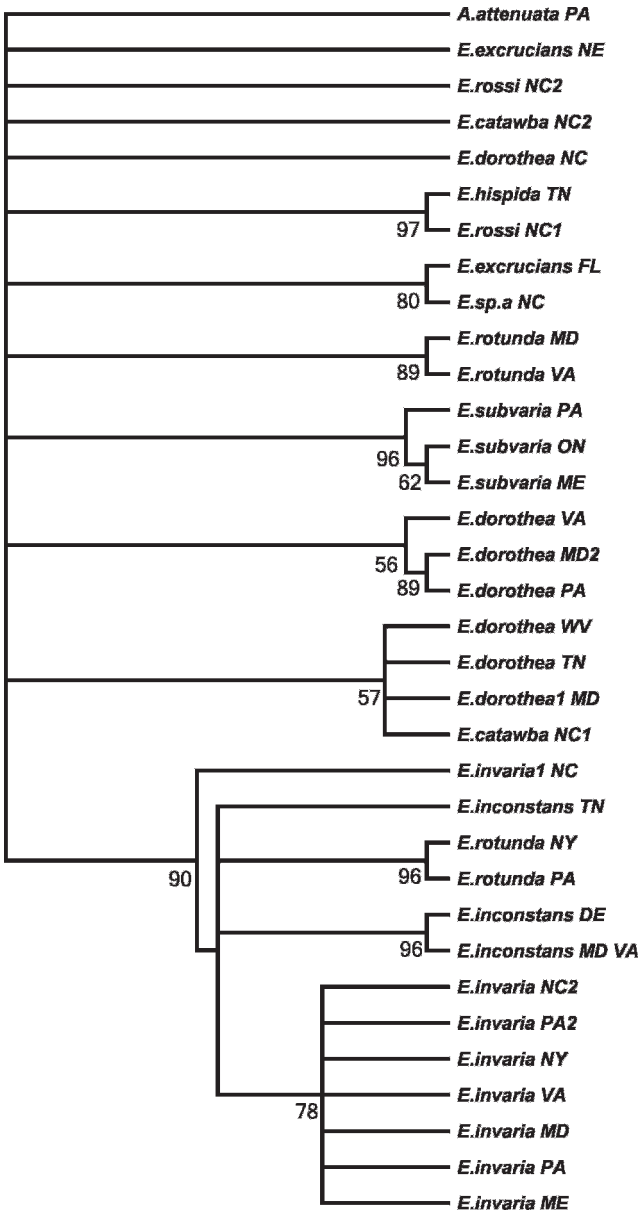

FIG. 3. Fifty percent majority-rule consensus neighbor-joining tree. Values shown are bootstrap support (1000 bootstrap replicates). *E.* = *Ephemerella* except *E. temporalis* = *Eurylophella temporalis*; *A. attenuata* = *Attenella attenuata*, *D. simplex* = *Dannella simplex*.

Distance-based barcoding had mixed results in our test assemblage of *Ephemerella* mayflies, which included multiple geographically distant populations from closely related species. The utility of mtDNA as a tool for rapid identification of species is based on the assumption that variation within species is typically an order of magnitude lower than that between species, creating a *barcoding gap* (Fig. 4B). However, the gap might be absent in emerging or closely related species because of incomplete lineage sorting (Funk and Omland 2003, Meyer and Paulay 2005, Meier et al. 2006). In this case, intraspecific variation overlaps with interspecific divergence and gives rise to genetically polyphyletic or paraphyletic species. When such overlap exists, the DNA marker (e.g., a DNA barcode) that is still in the process of sorting lineages cannot distinguish reliably among them. Our results (Fig. 4A) are consistent with the ''alternative version of the world with significant overlap and no gap'' described by Meyer and Paulay (2005) who, like Moritz and Cicero (2004), predicted that overlap of inter- and intraspecific variation would be greater when a larger proportion of closely related taxa are included, especially in taxonomically understudied groups. The average sequence length of *Ephemerella* samples used in our study was 418 bp, less than that recommended by the CBOL protocols (500–650 bp). However, shorter COI sequences from both the 5′ and 3′ halves of the standard mtDNA COI barcoding region were found to be comparable in sequence divergence by Hebert et al. (2003b), and COI barcoding standards from fragments of $\leq$150 bp have been developed (Hajibabaei et al. 2006, Meusnier et al. 2008). Our results indicate that the barcoding sequences and methods included enough phylogenetic information to identify successfully the lineages recovered from the maximum parsimony and maximum likelihood analyses. However, ~22% of the pairwise distances in our test assemblage fell into the overlap zone (Fig. 4A), and multiple comparisons within 4 species (*E. rossi*, *E. dorothea*, *E. catawba*, and *E. excrucians*) were diverged by >10%. We did not attempt to identify diagnostic nucleotide substitutions in the COI sequence as character states to separate taxa, as in Rach et al. (2008), but it is clear that genetic distance thresholds alone were not useful in identifying morphologically indistinct species in the *Ephemerella* test assemblage.

The distance data and sampling scale were useful in establishing geographic distributions of mtDNA lineages in *E. invaria*, *E. dorothea*, *E. subvaria*, and *E. rotunda*, which have large overlapping ranges in the eastern US. The sampled populations of *E. invaria* in New York and Pennsylvania have high levels of genetic diversity at multiple nuclear loci (DHF, unpublished data), evidence that historic long-range dispersal and current levels of gene flow are sufficient to maintain intraspecific diversity in these taxa. On the other hand, a monomorphic haplotype (K2P divergence <0.005%) and fixation at allozyme loci (LCA, unpublished data) in all populations identified as *E. inconstans* sampled over a 300-km distance in Maryland and Virginia are evidence of a population

TABLE 3. Kimura 2-parameter genetic distances within and among lineages in the INVARIA II clade. Values on the diagonal are distances within lineages; distances below diagonal are pairwise distances among lineages in the clade. Mean within-lineage distance is 1.3 ± 0.4%. Mean distance among lineages (off-diagonal values) is 5.2 ± 0.3%.

Number	Branch name	1	2	3	4	5
1	INVARIA I	1.3				
2	ROTUNDA I	4.1	0.4			
3	INCONSTANS	5.1	5.4	2.3		
4	*E. invaria* NC 1	5.3	5.3	6.5	1.3	
5	*E. inconstans* TN	4.7	4.0	5.7	6.1	–

bottleneck in this region. Mapping the distribution of genetic diversity in aquatic organisms is increasingly important in conservation and biomonitoring programs that use phylogenetic and population genetic data to protect and manage freshwater resources (Waples et al. 2001, Schwartz et al. 2007).

The genus *Ephemerella* has been a haven for morphologically undifferentiated species. Barcoding promises to ease greatly the identification of morphologically indistinct species, but the process assumes that species limits have been defined previously with multiple data sources, and that species thus described can be diagnosed accurately from the chosen barcoding gene sequence. High-throughput DNA barcoding methods have accelerated and improved the process

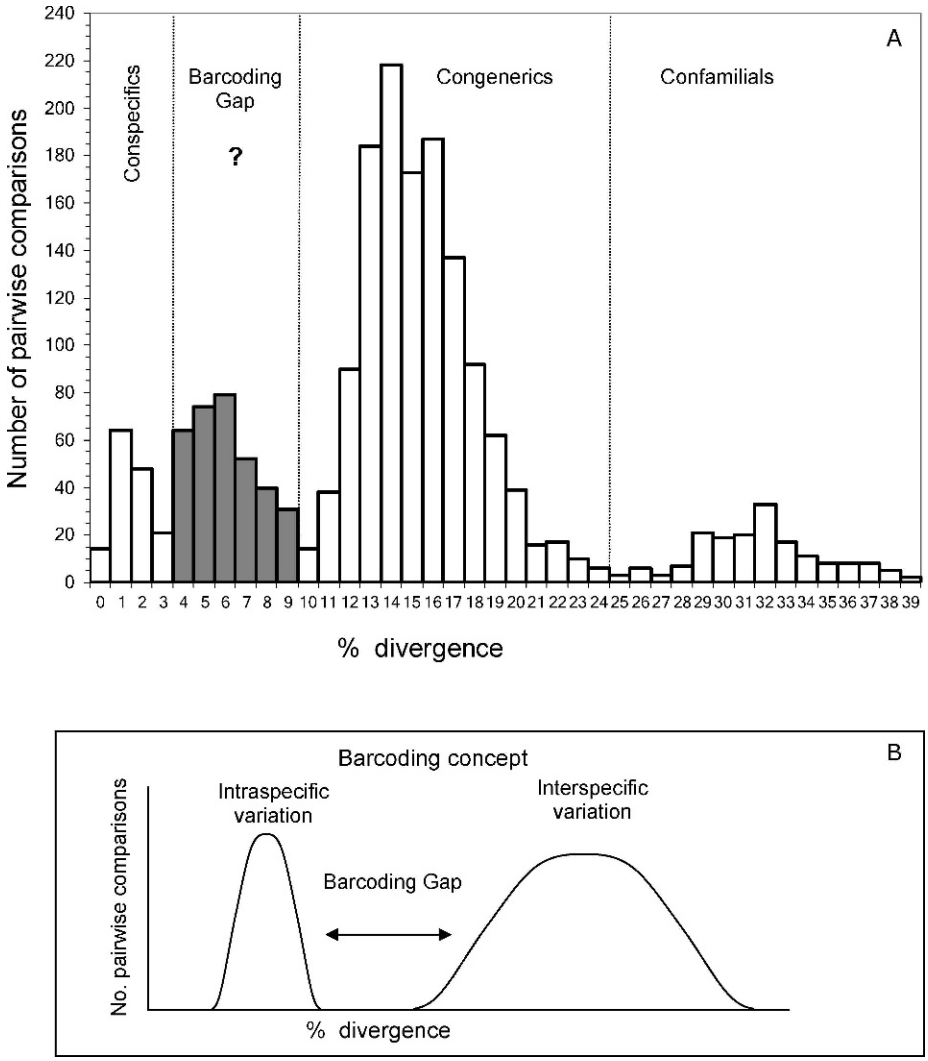

FIG. 4. Frequency distribution of pairwise genetic distance estimates. A.—The actual distribution of distances in the test assemblage. Considerable overlap between conspecifics and congenerics fills the barcoding gap. B.—The DNA barcoding concept predicts that intraspecific and interspecific genetic distances will be distributed bimodally, with a gap between conspecifics and congenerics. Modified from Meyer and Paulay (2005).

in many well-defined taxa and have added to our understanding of larval taxonomy and global aquatic insect diversity. However, until a more comprehensive systematic review is conducted and improved characters are discovered for defining and distinguishing species, attempts at DNA barcoding identification will encounter the same problems and reproduce the same errors that those of us using traditional methods have experienced with this group. Other widespread, genetically diverse, and morphologically ambiguous aquatic taxa exist (e.g., Trichoptera:Hydropsychidae:*Cheumatopsyche*, Plecoptera:Perlodidae:*Isoperla*) and accurate identification, especially of the larval forms, is a continuing challenge. Collaborative efforts in systematics, integrated taxonomy, and molecular barcoding (Trichoptera BOL 2007) are needed to meet the goal of accurate identification of aquatic insects using mtDNA.

Acknowledgements

We thank Ed DeWalt, C. J. Geraci, Dave Lenat, Bill Crouch, and 2 anonymous referees for comments that improved the manuscript; Jeff Shultz, Isaac Winkler, Andrea Badgley, and Joan West for technical assistance and advice; Luke Jacobus for preserved tissue samples; the Stroud Water Research Center and the North Carolina Division of Water Quality for fresh and preserved specimens. We also thank Dave Lenat for fresh specimens and information about the biology and ecology of *E. floripara*. This project was funded in part by the University of Maryland's Center for Biodiversity and Department of Entomology's Gahan Fellowship.

Literature Cited

ALLEN, R. K. 1980. Geographic distribution and reclassification of the subfamily Ephemerellinae (Ephemeroptera: Ephemerellidae). Pages 71–91 *in* J. F. Flannagan and K. E. Marshall (editors). Advances in Ephemeroptera biology. Plenum, New York.

ALLEN, R. K. 1984. A new classification of the subfamily Ephemerellinae and the description of a new genus. Pan-Pacific Entomologist 60:245–247.

ALLEN, R. K., AND G. F. EDMUNDS. 1962. A revision of the genus *Ephemerella* (Ephemeroptera: Ephemerellidae). V. The subgenus *Drunella* in North America. Miscellaneous Publications of the Entomological Society of America 4: 145–179.

ALLEN, R. K., AND G. F. EDMUNDS. 1963. A revision of the genus *Ephemerella* (Ephemeroptera: Ephemerellidae). VI. The subgenus *Serratella* in North America. Annals of the Entomological Society of America 56:583–600.

ALLEN, R. K., AND G. F. EDMUNDS. 1965. A revision of the genus *Ephemerella* (Ephemeroptera, Ephemerellidae). VIII. The subgenus *Ephemerella* in North America. Miscellaneous Publications of the Entomological Society of America 4:244–282.

ARMSTRONG, K. F., AND S. L. BALL. 2005. DNA barcodes for biosecurity: invasive species identification. Philosophical Transactions of the Royal Society Series B: Biological Sciences 360:1813–1823.

AVISE, J. C. 2004. Molecular markers, natural history, and evolution. 2nd edition. Sinauer, Sunderland, Massachusetts.

BALL, S. L., P. D. N. HEBERT, S. K. BURIAN, AND J. M. WEBB. 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. Journal of the North American Benthological Society 24:508–524.

BALLARD, J. W. O., AND M. C. WHITLOCK. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729–744.

BEKETOV, M. A. 2004. Different sensitivity of mayflies (Insecta, Ephemeroptera) to ammonia, nitrite and nitrate: linkage between experimental and observational data. Hydrobiologia 528:209–216.

BENKE, A. C., AND D. I. JACOBI. 1994. Production dynamics and resource utilization of snag-dwelling mayflies in a blackwater river. Ecology 75:1219–1232.

CBOL (CONSORTIUM FOR THE BARCODE OF LIFE). 2003. Consortium for the Barcode of Life. National Museum of Natural History, Washington, DC. (Available from: http://barcoding.si.edu/)

FINN, D. S., M. S. BLOUIN, AND D. A. LYTLE. 2007. Population genetic structure reveals terrestrial affinities for a headwater stream insect. Freshwater Biology 52: 1881–1897.

FUNK, D. J., AND K. E. OMLAND. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecological and Evolutionary Systematics 34: 397–423.

HAJIBABAEI, M., M. A. SMITH, D. H. JANZEN, J. J. RODRIGUEZ, J. B. WHITFIELD, AND P. D. N. HEBERT. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6:959–964.

HALL, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analyses program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95–98.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160–174.

HEBERT, P. D. N., A. CYWINSKA, S. L. BALL, AND J. R. DEWAARD. 2003a. Biological identifications through DNA barcodes. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 270:313–321.

HEBERT, P. D. N., S. RATNASINGHAM, AND J. R. DEWAARD. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences (Supplement) 270:S96–S99.

HILSENHOFF, W. L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomologist 20:31–39.

iBOL (International Barcode of Life). 2003. International barcode of life. International Consortium Initiative, Biodiversity Institute of Ontario, Ontario. (Available from: http://www.dnabarcoding.org)

Jacobus, L. M., and W. P. McCafferty. 2003. Revisionary contributions to North American *Ephemerella* and *Serratella* (Ephemeroptera: Ephemerellidae). Journal of the New York Entomological Society 111:174–193.

Jacobus, L. M., and W. P. McCafferty. 2008. Revision of Ephemerellidae genera (Ephemeroptera). Transactions of the American Entomological Society 134:185–274.

Lenat, D. R. 1993. A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water quality ratings. Journal of the North American Benthological Society 12:279–290.

Lytle, D. A., M. T. Bogan, and D. S. Finn. 2008. Evolution of aquatic insect behaviours across a gradient of disturbance predictability. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 275:453–462.

McCafferty, W. P., and T. Q. Wang. 2000. Phylogenetic systematics of the major lineages of pannote mayflies (Ephemeroptera: Pannota). Transactions of the American Entomological Society 126:9–101.

McShaffrey, D., and W. P. McCafferty. 1991. Ecological association of the mayfly *Ephemerella needhami* (Ephemeroptera: Ephemerellidae) and the green alga *Cladophora* (Chlorophyta: Cladophoraceae). Journal of Freshwater Ecology 6:383–394.

Meier, R., K. Shiyang, G. Vaidya, and P. K. C. Ng. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55:715–728.

Meusnier, I., G. A. C. Singer, J. F. Landry, D. A. Hickey, P. D. N. Hebert, and M. Hajibabaei. 2008. A universal DNA mini-barcode for biodiversity analysis. BioMed Central Genomics 9:214.

Meyer, C. P., and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. PloS Biology 3: e422. doi:10.1371/journal.pbio.0030422.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.

Moritz, C., and C. Cicero. 2004. DNA barcoding: promise & pitfalls. PLoS Biology 2:e354. doi:10.1371/journal.pbio.0020354.

Nelson, L. A., J. F. Wallman, and M. Dowton. 2007. Using COI barcodes to identify forensically and medically important blowflies. Medical and Veterinary Entomology 21:44–52.

Pfenninger, M., C. Nowak, C. Kley, D. Steinke, and B. Streit. 2007. Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. Molecular Ecology 16:1957–1968.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Rach, J., R. DeSalle, I. N. Sarkar, B. Schierwater, and H. Hadrys. 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Proceedings of the Royal Society B: Biological Sciences 275:237–247.

Ratnasingham, S., and P. D. N. Hebert. 2007. BOLD: the barcode of life data system. Molecular Ecology Notes 7: 355–364.

Rezanka, K. M., and A. E. Hershey. 2003. Examining primary producer–consumer interactions in a Lake Superior tributary using ^{15}N-tracer, grazer-reduction, and nutrient-bioassay experiments. Journal of the North American Benthological Society 22:371–387.

Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. Conservation Biology 20:1026–1033.

Schwartz, M. K., G. Luikart, and R. S. Waples. 2007. Genetic monitoring: a promising tool for conservation and management. Trends in Ecology and Evolution 22:25–33.

Shaw, K. L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences of the United States of America 99:16122–16127.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 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.

Smith, P. J., S. M. McVeagh, and D. Steinke. 2008. DNA barcoding for the identification of smoked fish products. Journal of Fish Biology 72:464–471.

Swofford, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.09b. Sinauer, Sunderland, Massachusetts.

Trichoptera BOL (Trichoptera Barcode of Life). 2007. Trichoptera Barcode of Life website. University of Guelph, Guelph, Ontario. (Available from: http://www.trichopterabol.org)

Waples, R. S., R. G. Gustafson, L. A. Weitkamp, J. M. Myers, O. W. Johnson, P. J. Busby, J. J. Hard, G. J. Bryant, F. W. Waknitz, K. Nelly, D. Teel, W. S. Grant, G. A. Winans, S. Phelps, A. Marshall, and B. M. Baker. 2001. Characterizing diversity in salmon from the Pacific Northwest. Journal of Fish Biology 59:1–41.

Will, K. W., and D. Rubinoff. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD dissertation, University of Texas at Austin, Austin, Texas.

Received: 20 October 2008
Accepted: 10 April 2009