

Phylogeographic and nested clade analysis of the stonefly *Pteronarcys californica* (Plecoptera:Pteronarcyidae) in the western USA

JOHN S. K. KAUWE¹

Department of Biology, Washington University, St. Louis, Missouri 63110 USA

DENNIS K. SHIOZAWA²

Department of Integrative Biology, Brigham Young University, Provo, Utah 84602 USA

R. PAUL EVANS³

Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah 84602 USA

Abstract. Long-distance dispersal by aquatic insects can be difficult to detect because direct measurement methods are expensive and inefficient. When dispersal results in gene flow, signs of that dispersal can be detected in the pattern of genetic variation within and between populations. Four hundred seventy-five base pairs of the mitochondrial gene, cytochrome *b*, were examined to investigate the pattern of genetic variation in populations of the stonefly *Pteronarcys californica* and to determine if long-distance dispersal could have contributed to this pattern. Sequences were obtained from 235 individuals from 31 different populations in the western United States. Sequences also were obtained for *Pteronarcella badia*, *Pteronarcys dorsata*, *Pteronarcys princeps*, *Pteronarcys proteus*, and *Pteronarcys biloba*. Phylogenies were constructed using all of the samples. Nested clade analysis on the *P. californica* sequence data was used to infer the processes that have generated the observed patterns of genetic variation. An eastern North American origin and 2 distinct genetic lineages of *P. californica* could be inferred from the analysis. Most of the current population structure in both lineages was explained by a pattern of restricted gene flow with isolation by distance (presumably the result of dispersal via connected streams and rivers), but our analyses also suggested that long-distance, overland dispersal has contributed to the observed pattern of genetic variation.

Key words: phylogeography, nested clade analysis, Plecoptera, dispersal, western United States.

Studies investigating the dispersal of aquatic insects have become common, but much remains to be learned about patterns of dispersal and mechanisms of gene flow (Bilton et al. 2001). Dispersal poses risks for any organism, but it also can be beneficial (Stenseth and Lidicker 1992). It allows organisms to colonize new and open habitat, and it reduces the risk of inbreeding depression. Dispersal that results in gene flow between populations increases genetic variation within a population and decreases genetic variation between populations. Conversely, absence of gene flow between populations can result in evolutionarily independent populations of the same species (Bilton et al. 2001).

Aquatic invertebrates disperse via active and passive mechanisms (Bilton et al. 2001). Active dispersal is accomplished mostly by adult flight.

Passive dispersal is often the result of downstream drift (Waters 1972), but it also can occur when insects are carried to new habitat by animal vectors or wind (Maguire 1963). Even insects that are weak fliers can experience long-distance dispersal when aided by wind (Kelly et al. 2001). Invertebrates living in riverine environments have additional opportunities for passive dispersal. Insects can cross drainage divides from stream headwaters if they actively disperse upstream as adults after drifting passively downstream as larvae or nymphs (Müller 1954). Historical processes such as stream-and-river-capture events also can allow aquatic organisms to disperse into previously isolated drainages (Taylor 1985).

Many mechanisms of long-distance dispersal are known, but the extent to which each mechanism actually occurs is unclear. Direct estimates of overland dispersal by aquatic invertebrates involve the use of various types of traps (Jackson and Resh 1989, Kovats et al. 1996, Col-

¹ E-mail addresses: keoni@icarus.wustl.edu

² shiozawa@byu.edu

³ evansp@byu.edu

lier and Smith 1998, Griffith et al. 1998), radioactive and stable isotopes (Service 1993), or transponder tags (Riley et al. 1996). These tools provide useful information about present-day dispersal, but they are unlikely to detect rare chance events or dispersal associated with the geologic history of a region. Such events can be assessed with indirect measures of dispersal.

These indirect measures rely on the fact that dispersal resulting in gene flow can alter the spatial pattern of genetic variation in the species. Thus, dispersal can be detected through careful analysis of genetic information. Allozymes, mitochondrial DNA (mtDNA) haplotypes, randomly amplified polymorphic DNA, and microsatellites have been used to investigate gene flow in aquatic invertebrates (e.g., Crease et al. 1997, Taylor et al. 1998, Thomas et al. 1998, Hughes et al. 1999, Freeland et al. 2000, Monaghan et al. 2001, Schultheis et al. 2002). Population genetic parameters also can be estimated from these genetic markers (Wright 1951, Nei 1972). A relatively new approach, nested clade analysis (NCA), provides a unique and especially sensitive means by which geographical data and allele frequencies can be used to differentiate between recurring and historical dispersal events (Templeton 1998).

NCA uses a haplotype network to evaluate the relationship between mtDNA haplotypes and geography. NCA assumes that haplotypes and clades on the interior of the network are ancestral and that haplotypes on the tips of the network are more recent. This temporal information is then used to partition the recent and historical processes responsible for the observed pattern of genetic variation. The method tries to reject the null hypothesis of no association between haplotype variation and geography. If the null hypothesis is rejected, then an explicit inference key is used to interpret the statistically significant associations. This method allows inference of multiple, separate events in space and time (e.g., fragmentation, range expansion, gene flow, and genetic drift; Templeton 1998, Pfenninger and Posada 2002).

Various hypotheses have been generated to explain stonefly dispersal. Predation on stonefly eggs by stream trout, and the subsequent dispersal of eggs that have passed through the digestive system unharmed, may be a dispersal mechanism (Hartleb and Timm 2000). Recent studies suggest that movement by nymphs is an

important mechanism of dispersal for the stonefly *Peltoperla tarteri* (Schultheis et al. 2002). The main method of invasion of new habitat by stoneflies may be overland dispersal (Hughes et al. 1999, Ketmaier et al. 2001), but other studies suggest that, in some cases, overland dispersal may not be as important as historical geography (Houseman and Baumann 1997, Baumann et al. 1999).

Pteronarcys californica Newport is well-known among the species of stoneflies found in the western United States, and it has a widespread distribution. *Pteronarcys californica* is the largest of all stoneflies. The nymphs are usually dark brown and average between 35 and 50 mm in length. *Pteronarcys californica* usually is found in 2nd- to 7th-order streams with swift, clean water and rocky bottoms. These stoneflies are shredders (Merritt and Cummins 1996), and their distribution is patchy; densities are high in areas with large amounts of allochthonous input and very low in areas with little or no allochthonous input (DKS, unpublished data). The nymphs undergo multiple instars as they progress through a 2- to 4-y life cycle (Stewart and Stark 1993). Very little is known about the dispersal mechanisms and capabilities of *P. californica*. However, its large size and flight capabilities suggest that overland movement by adults may be a significant mechanism of dispersal in this organism.

Our study used the mitochondrial gene, cytochrome *b* (*cyt b*), from populations of *P. californica* to determine whether long-distance, overland dispersal resulting in gene flow has been a factor in shaping patterns of genetic variation in this species. A haplotype network was used to describe the pattern of genetic variation in *P. californica*. Significant events in the dispersal history of *P. californica* were then inferred by statistical analyses of that pattern.

Methods

Sampling

Pteronarcys californica nymphs were collected from 31 sites in the western US (Fig. 1). *Pteronarcys princeps* nymphs were collected from several western states, and *Pteronarcella badia* nymphs were collected from Utah. *Pteronarcys dorsata* nymphs were collected from sites in northern Minnesota, Alabama, and North Carolina. *Pteronarcys proteus* nymphs were collected

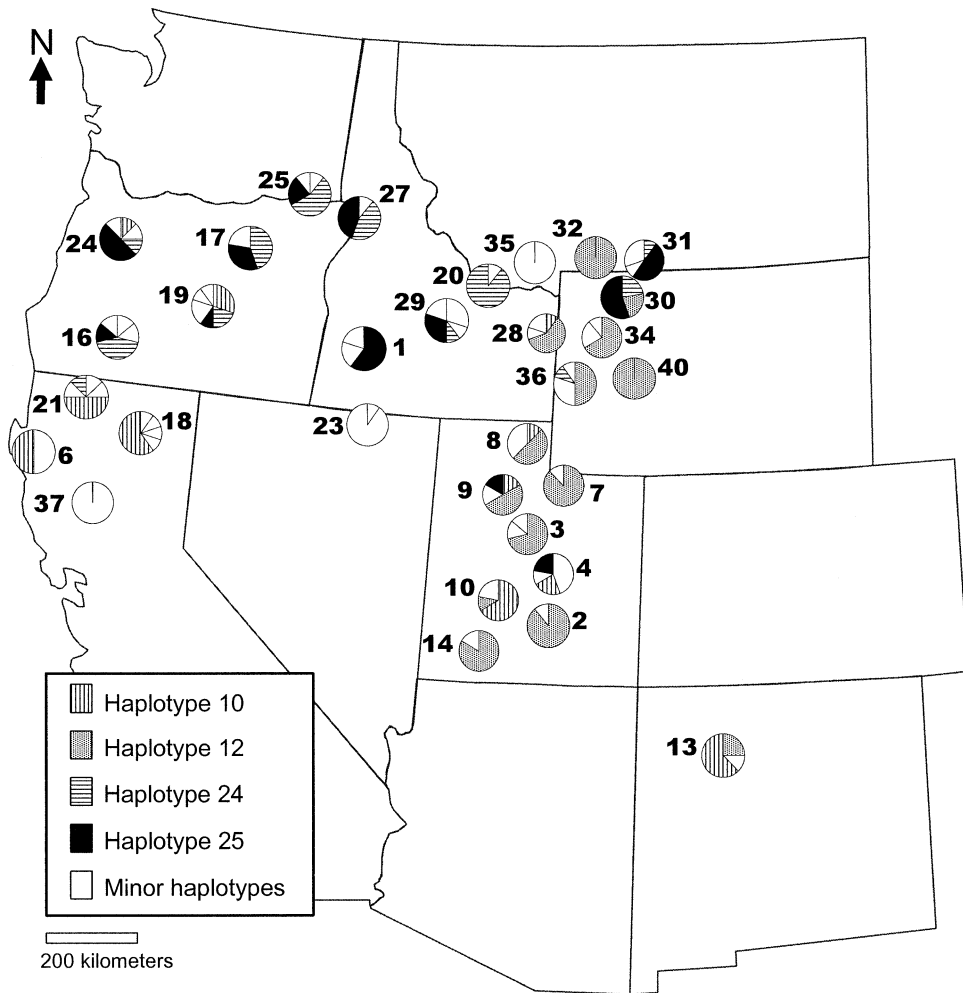


FIG. 1. The distribution of *Pteronarcys californica* sampling sites in western USA. Pie diagrams show the % composition of haplotypes at each site. Population identification numbers are as in Table 1. Major haplotypes (frequency > 0.1; see Table 2) are identified by patterns (inset); minor haplotypes (frequency < 0.1) are in white.

from sites in Pennsylvania and North Carolina, and *Pteronarcys biloba* nymphs were collected from sites in North Carolina. All specimens were preserved in 95% ethanol immediately after collection. Global positioning system (GPS) coordinates were recorded at each site (*P. californica* sample sites can be found in Table 1). Voucher specimens were identified by Drs Riley Nelson and Richard Baumann, Brigham Young University.

DNA was extracted from up to 10 individuals from each population of *P. californica* ($n = 235$ individuals) using the DNeasy protocol (Qia-

gen, Valencia, California). DNA was extracted from 11 *P. princeps*, 9 *P. proteus*, 5 *P. dorsata*, and 3 *P. biloba* using the same method. Polymerase chain reaction (PCR) was performed using 2 stonefly-specific primers, Plecoptera *cyt-b* F (5'-TGT CCA TAT TTG YCG AGA TGT-3') and Plecoptera *cyt-b* R (5'-CTT ATG TTT TCA AAA CAT ATG C-3'). These primers were designed by examining other insect *cyt b* sequences and targeting conserved regions. The primers amplified ~850 base pairs (bp) of the mitochondrial *cyt b* gene. PCR was performed in 50- μ L reaction mixtures consisting of DNA template

TABLE 1. Population identification (ID) number, number of mitochondrial DNA (mtDNA) sequences sampled, and location for each population of *Pteronarcys californica*. GPS = global positioning system.

ID	mtDNA sequences	River or stream	State/county	GPS coordinates (Latitude longitude)
1	5	South Fork Boise River	Idaho/Elmore	43°34.00 N 115°83.00 W
2	9	Huntington Creek	Utah/Emory	39°46.00 N 111°20.00 W
3	7	Provo River	Utah/Utah	40°20.00 N 111°35.00 W
4	6	Diamond Fork	Utah/Utah	40°04.00 N 111°30.00 W
6	2	Russian River	California/Mendocino	39°20.00 N 123°60.00 W
7	8	Ogden River	Utah/Weber	41°17.00 N 111°40.00 W
8	8	Blacksmith Fork	Utah/Cache	41°35.00 N 111°50.00 W
9	6	Ogden River	Utah/Weber	41°17.00 N 112°00.00 W
10	9	Beaver River	Utah/Beaver	38°12.00 N 112°52.30 W
13	8	Pecos River	New Mexico/San Miguel	35°10.07 N 105°06.48 W
14	6	Mammoth Creek	Utah/Garfield	37°38.00 N 112°30.00 W
16	7	Cow Creek	Oregon/Jackson	42°54.84 N 123°28.78 W
17	9	Middle Fork John Day River	Oregon/Grant	44°37.31 N 118°34.82 W
18	10	Pitt River	California/Shasta	41°00.27 N 121°43.22 W
19	10	Canyon Creek	Oregon/Grant	44°15.39 N 118°55.44 W
20	9	Birch Creek	Idaho/Clark	44°80.35 N 112°54.02 W
21	9	Beaver Creek	California/Siskiyou	41°52.23 N 122°49.00 W
23	10	Upper Marys River	Nevada/Elko	41°24.00 N 115°13.00 W
24	8	North Fork Alsea River	Oregon/Benton	44°24.95 N 123°33.73 W
25	9	Mill Creek	Washington/Walla Walla	46°50.13 N 118°13.69 W
27	9	McKinley Creek	Idaho/Valley	44°40.00 N 116°15.00 W
28	10	Fall River	Idaho/Fremont	44°01.27 N 111°75.23 W
29	10	Salmon River	Idaho/Custer	44°28.23 N 114°12.03 W
30	9	Clark's Fork	Wyoming/Park	44°50.51 N 109°19.04 W
31	10	Tongue River	Wyoming/Sheridan	44°52.69 N 107°16.09 W
32	10	Yellowstone River	Montana/Park	45°06.18 N 110°47.44 W
34	9	Hoback River	Wyoming	43°19.03 N 110°42.03 W
35	1	Madison River	Montana/Madison	44°49.60 N 111°39.24 W
36	10	Grays River	Wyoming/Lincoln	43°08.73 N 110°58.26 W
37	1	North Fork Cache Creek	California/Lake	38°98.83 N 122°53.88 W
40	1	Pine Creek	Wyoming/Pine Dale	42°52.00 N 109°52.00 W

(100 ng), deoxyribonucleotides (0.125 mM each), primers (10 pmoles each), buffer (10 mM Tris-HCl; 1.5 mM MgCl₂, 25 mM KCl), and Taq polymerase (0.5 units) as follows: 4 min at 95°C; 34 cycles of 20 s at 94°C, 30 s at 47°C, and 1.5 min at 72°C; and 7 min at 72°C. Purification of the PCR product was performed using the QBiogene Gene Clean III protocol (Qbiogene, Carlsbad, California). Cycle sequencing was performed using the ABI Big Dye terminator protocol (Applied Biosystems, Foster City, California). Samples were submitted to the Brigham Young University DNA Sequencing Center and were sequenced on an ABI 377 automated sequencer.

Sequence alignment

Sequences were edited and aligned with Sequencher 3.0 software (Gene Codes Corporation,

Ann Arbor, Michigan). The alignment was unequivocal because the sequences included no insertions, deletions, or repeats. Sequences were trimmed to 475 bp.

Phylogenetic analyses

The 235 *P. californica* sequences were collapsed into haplotypes. Haplotypes with frequencies <0.01 were derivatives of the more common haplotypes and were excluded from the phylogenetic analysis because of limits on computational power. Thus, the analysis included 10 *P. californica* haplotypes and the sequence data from 5 other species (*P. biloba*, *P. dorsata*, *P. princeps*, *P. proteus*, and *Pteronarcella badia*). Sequences from *Pteronarcella badia* were used to root the tree. The phylogeny was estimated using the

maximum parsimony (MP) setting in MEGA 2.0 (Kumar et al. 2001). A set of MP trees was obtained using close-neighbor-interchange (search level = 2, random additions = 200). Bootstrap consensus values were calculated from 225 replicates. Mean genetic diversities were calculated using the Kimura 2-parameter model in MEGA 2.0.

Nested clade and GeoDis analyses

A network of *P. californica* haplotypes was constructed using statistical parsimony as implemented by the software program TCS (Clement et al. 2000). Ambiguities in the network (parts of the network where a single haplotype is connected to >2 other haplotypes) were resolved according to 3 criteria derived from coalescent theory: 1) geography—haplotypes are more likely to be connected to haplotypes from the same population or region than to haplotypes occurring in distant populations; 2) topology—haplotypes are more likely to be connected to interior (ancestral) haplotypes than to tip (more recent) haplotypes; and 3) frequency—haplotypes are more likely to be connected to haplotypes with high frequency than to singletons (for further detail see Pfenninger and Posada 2002). NCA (Templeton 1998), an objective method for assessing associations of geography and haplotypes (Cruzan and Templeton 2000), was used to determine the nesting structure of the network.

Other methods exist for using haplotype data to investigate population history (mismatch pair distribution, Rogers and Harpending 1992; the skyline plot method, Pybus et al. 2000; and statistical phylogeography, Knowles and Maddison 2002). However, these other methods investigate single historical events, deal exclusively with expansion and contraction of populations, or require specific, a priori hypotheses of historical events. In contrast, NCA can help reconstruct complex phylogeographical histories with little or no prior information (Fetzner and Crandall 2003, Templeton 2004). The software program GeoDis 2.0 (Posada et al. 2000; <http://darwin.uvigo.es/software/geodis.html>) was used to calculate the NCA distance measures and determine the statistical relationship between the haplotypes in the network and geography. This program calculates the following NCA distances: 1) average clade distance (D_c), the average

distance of all clade members from the geographical center of the clade relative to the other clades within the same nesting clade; 2) nested clade distance (D_n), a measure of how widespread a clade is relative to the distribution of its nesting clade; and 3) interior-tip (I-T) distances, measures that indicate how widespread younger clades (tip clades) are compared to their ancestors (interior clades) relative to other clades within the same nesting clade. Statistical significance of these distances was tested by comparing each value with a null distribution generated by random permutation of clades against sample sites (Pfenninger and Posada 2002). Euclidean distances between sites were based on GPS coordinates. Euclidean distances were used in the GeoDis calculations rather than river distances because stoneflies are capable of overland dispersal and specific hypotheses of ancient drainage connections do not exist (Fetzner and Crandall 2003). The associations that were significant at $\alpha = 0.05$ in the GeoDis calculations were interpreted using the updated version of the inference key (Templeton 1998; most recent update available from: http://darwin.uvigo.es/download/geodisKey_14Jul04.pdf).

Results

Phylogenetic analyses

Pteronarcys californica had 33 unique haplotypes (Table 2), but only 10 of these had a frequency >0.01. These were haplotypes 10, 12, 13, 17, 21, 23, 24, 25, 27, and 30 (Table 2). Haplotype 23 was basal to all other *P. californica* haplotypes (Fig. 2). Haplotypes 24, 25, 27, and 30 were basal to haplotypes 10, 12, 13, 17, and 21. A strict consensus (cutoff level = 75%) was used to summarize 1344 MP trees (Fig. 2). The basal nodes of this phylogeny were well-resolved. They supported previously published, morphology-based phylogenies in that they defined the eastern species, *P. dorsata*, *P. proteus*, and *P. biloba*, as basal to *P. californica* and *P. princeps* (Nelson 1988). However, our phylogeny placed specimens identified as *P. proteus* collected from Pennsylvania with specimens identified as *P. biloba* collected from North Carolina.

NCA and GeoDis analysis

The nested cladogram consisted of 3 levels: thirty-three 0-step clades (each individual hap-

TABLE 2. *Pteronarcys californica* haplotypes, the population in which each haplotype occurred, and the frequency of each haplotype with respect to the total number of individuals sampled. Population identification (ID) numbers are as in Table 1.

Haplotype	Population ID	Haplotype frequency
1	18	0.004
2	6	0.004
3	18	0.004
4	18	0.004
5	18	0.004
6	21	0.004
7	21	0.004
8	37	0.004
9	4	0.004
10	4, 6, 8, 9, 10, 13, 18, 19, 21, 24, 28	0.119
11	13	0.004
12	2, 3, 7, 8, 9, 10, 13, 14, 28, 30, 32, 34, 36, 40	0.289
13	9, 25, 29	0.021
14	28	0.004
15	21	0.004
16	10	0.008
17	3, 7, 8, 28, 34, 35, 36	0.055
18	2	0.004
19	3, 34	0.008
20	16, 24	0.008
21	4, 14, 16, 27, 29	0.021
22	23	0.004
23	20, 23	0.042
24	16, 17, 19, 20, 21, 24, 25, 27, 29, 30, 31, 36	0.144
25	1, 4, 9, 16, 17, 19, 24, 25, 27, 29, 30, 31	0.144
26	16	0.004
27	17, 19, 25, 29, 36	0.034
28	19	0.004
29	31	0.004
30	31	0.012
31	24	0.004
32	1, 19	0.008
33	1	0.004

lotype), six 1-step clades (haplotypes within 1 mutational step of each other; designated 1–1 through 1–6 in Fig. 3), and two 2-step clades (haplotypes within 2 mutational steps of each other; designated 2–1 and 2–2 in Fig. 3). Ambiguities existed in 5 different locations on the cladogram (A, B, C, D, and E in Fig. 3). A, C, and D were resolved by geographic criteria; B was resolved using frequency criteria; and E

was resolved using both geographic and frequency criteria (Pfenninger and Posada 2002).

In clade 1–1, the D_c of interior haplotype 10 ($D_c = 531.74$) and the I–T distance ($D_c = 455.04$) were significantly large (Table 3). In clade 1–2, the D_n of tip haplotype 15 ($D_n = 971.29$) was significantly large, and the D_c of tip haplotype 17 ($D_c = 151.07$) was significantly small. In clade 1–3, the D_c and the D_n of interior haplotype 24 ($D_c = 320.37$, $D_n = 333.35$) and the I–T distance ($D_c = 338.62$, $D_n = -301.91$) were significantly small. In clade 1–4, the D_c and the D_n of interior haplotype 23 ($D_c = 162.29$, $D_n = 152.68$) and the I–T distance ($D_c = 162.29$, $D_n = 65.73$) were significantly large. No significant associations were observed in clade 1–5. In clade 1–6, the D_c of tip haplotype 30 ($D_c = 0$) was significantly small, and the I–T distance ($D_c = 332.57$) was significantly large. The D_n for tip haplotypes 20 ($D_n = 667.58$), 30 ($D_n = 641.14$), and 31 ($D_n = 660.94$) were significantly large. In clade 2–1, the D_n for interior clade 1–1 ($D_n = 626.63$) and the D_c and D_n for I–T distance ($D_c = 213.04$, $D_n = 263.79$) were significantly large, and the D_c and D_n for tip clade 1–2 ($D_c = 274.59$, $D_n = 362.84$) were significantly small. In clade 2–2, the D_c for tip clades 1–4 ($D_c = 141.93$) and 1–5 ($D_c = 216.58$) and the D_n for tip clade 1–5 ($D_n = 248.69$) were all significantly small. In the total cladogram, the D_n for tip clade 2–1 ($D_n = 526.33$) was significantly small, and the D_n for interior clade 2–2 ($D_n = 565.42$) and the I–T distance ($D_n = 39.09$) were significantly large.

Discussion

Phylogenetic analyses

The family Pteronarcyidae includes 12 extant species, found only in North America. Ten of these are members of the genus *Pteronarcys*, and all but 2 of these 10 are found exclusively east of the continental divide. *Pteronarcys dorsata* is a sister taxon of the western clade, and its range extends to the eastern border of the continental divide (Nelson 1988, Kondratieff and Baumann 2000). *Pteronarcys californica* is found throughout the northern and eastern Basin and Range province of the western US but is conspicuously absent in the western Great Basin, where *P. princeps* is the sole representative of the genus (Sheldon 1979). Published phylogenies of the family Pteronarcyidae regard the species on the eastern

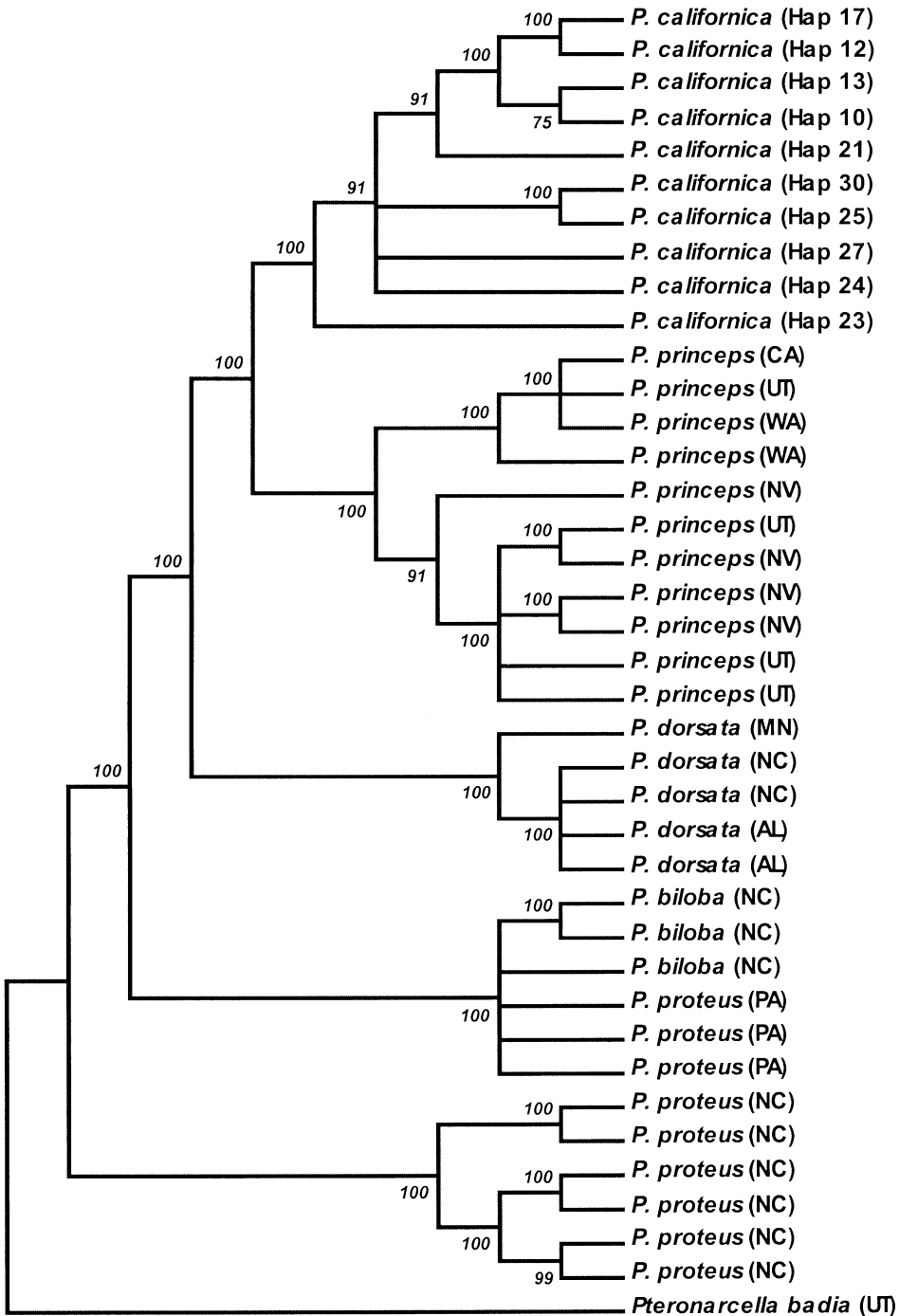


FIG. 2. Strict consensus (cutoff level = 75%) of 1344 maximum parsimony trees. Phylogeny defines the position of the *Pteronarcys californica*/*P. princeps* clade in the family Pteronarcyidae. *Pteronarcys californica* haplotype (Hap) identification numbers are in parentheses following each entry. See Table 1 for population details and Table 2 for haplotype identification numbers and details. States in which other species were collected are indicated in parentheses following each entry. State abbreviations are: Alabama (AL), California (CA), Montana (MN), North Carolina (NC), Nevada (NV), Pennsylvania (PA), Utah (UT), and Washington (WA).

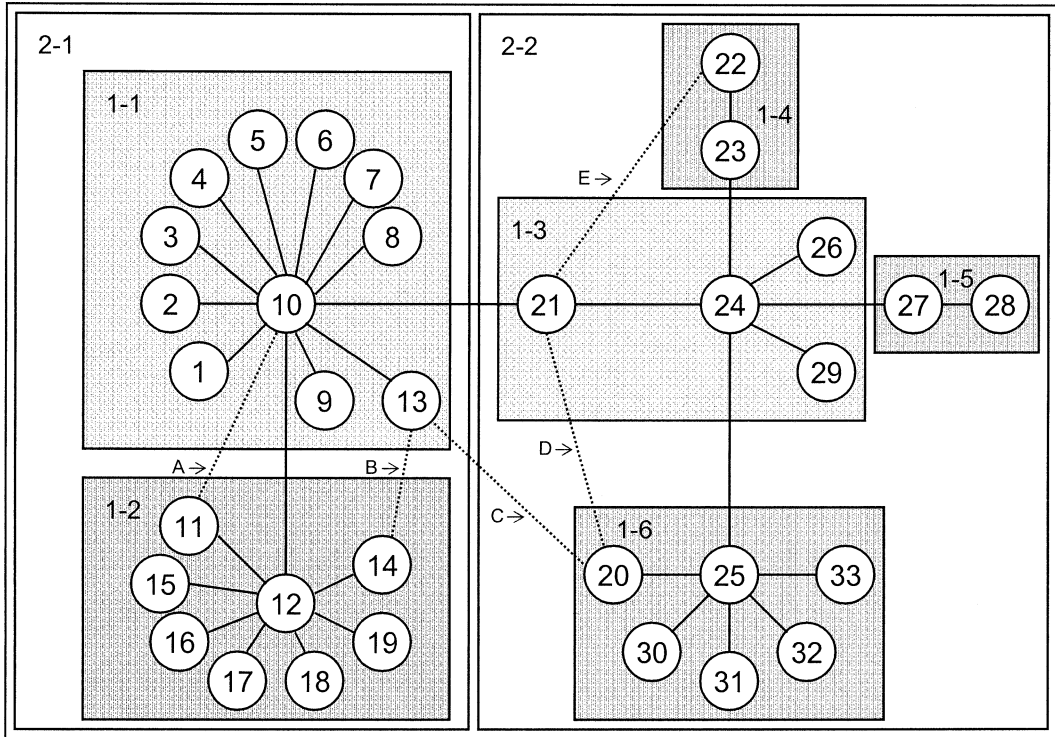


FIG. 3. Representation of the genetic structure of *Pteronarcys californica* based on nested clade analysis. Identification numbers (1–33) in open circles refer to haplotypes (0-step clades) described in Table 2. One-step clades in dark boxes are labeled 1–1 through 1–6. Two-step clades enclosed in light boxes are labeled 2–1 and 2–2. Letters (A, B, C, D, and E) refer to ambiguities in the network; see text for explanation. A solid line indicates a connection between haplotypes. A dashed line indicates a relationship that was broken to solve ambiguities in the network.

side of the continental divide as basal to the western clade of *P. californica*/*P. princeps* (Nelson 1988). The results of our phylogenetic analysis also support the hypothesis that the western clade is derived from an eastern ancestor (Fig. 2). Interestingly, specimens that were identified morphologically as *P. proteus* from Pennsylvania clearly were included in a clade with *P. biloba* from North Carolina. This group was distinct from the North Carolina *P. proteus* clade (Fig. 2), which was basal to all other *Pteronarcys* included in this analysis. This paraphyletic association in *P. proteus* could have been the result of contamination of DNA in some step of the analysis, a lack of consistency in the morphological characters used to identify nymphs, or introgression between the 2 species. Introgression could have caused mtDNA from *P. proteus* to become fixed in *P. biloba* in the northern part of its range. The possibility of DNA contamination was evaluated

by isolating DNA from additional specimens from Pennsylvania collections. The additional analyses generated the same phylogenetic relationships, supporting the close phyletic association of Pennsylvania *P. proteus* with North Carolina *P. biloba*. In addition, the specimens were re-examined, and they clearly fit the key morphological characters for *P. proteus* as given in Ricker (1952). Therefore, the most likely explanations for the grouping are that introgression has occurred or that nymphal characters are more variable than has been assumed in the literature. This result was not part of the main objective of this study, but it certainly warrants further investigation.

NCA inferences

The results of the NCA suggest that the historical events that led to the present-day popu-

TABLE 3. Nested clade analysis of the haplotype network with average clade distance (D_c) and nested clade distance (D_n) for each haplotype and clade, and interior–tip (I–T) distances for each clade. Distances that were significantly large or small (via permutation tests) are labeled (^l) and (^s), respectively.

	Topology	Haplotype	D_c	D_n
Clade 1-1	Tip	1	0	257.51
	Tip	2	0	467.05
	Tip	3	0	257.51
	Tip	4	0	257.51
	Tip	5	0	257.51
	Tip	6	0	373.44
	Tip	7	0	373.44
	Tip	8	0	394.18
	Tip	9	0	614.85
	Interior	10	531.74 ^l	525.39
Clade 1-1	Tip	13	214.77	599.58
	I–T		455.04 ^l	78.89
Clade 1-2	Tip	11	0	890.65
	Interior	12	272.55	267.79
	Tip	14	0	265.52
	Tip	15	0	971.29 ^l
	Tip	16	0	416.06
	Tip	17	151.07 ^s	232.53
	Tip	18	0	214.14
	Tip	19	167.09	168.2
Clade 1-3	Interior	21	459.07	500.86
	Interior	24	320.37 ^s	333.35 ^s
	Tip	26	0	585.17
	Tip	29	0	729.42
Clade 1-3	I–T		338.62 ^s	–301.91 ^s
	Clade 1-4	Tip	22	0
Interior		23	162.29 ^l	152.68 ^l
Clade 1-4	I–T		162.29 ^l	65.73 ^l
	Clade 1-5	Interior	27	226.53
Tip		28	0	164.85
Clade 1-5	I–T		226.53	57.94
	Clade 1-6	Tip	20	83.18
Interior		25	378.21	376.23
Tip		30	0 ^s	641.14 ^l
Tip		31	0	660.94 ^l
Tip		32	122.23	138.47
Tip		33	0	61.79
Clade 1-6	I–T		332.57 ^l	–96.9
	Clade 2-1	Interior	Clade 1-1	487.63
Tip		Clade 1-2	274.59 ^s	362.84 ^s
Clade 2-1	I–T		213.04 ^l	263.79 ^l
	Clade 2-2	Interior	Clade 1-3	374.23
Tip		Clade 1-4	141.93 ^s	296.27
Tip		Clade 1-5	216.58 ^s	248.69 ^s
Tip		Clade 1-6	387.16	386.32
I–T			38.91	25.9
Cladogram	Tip	Clade 2-1	528.92	526.33 ^s
	Interior	Clade 2-2	551.96	565.42 ^l
Cladogram	I–T		23.04	39.09 ^l

TABLE 4. The inference chain and inference(s) for each clade (inference key is available at http://darwin.uvigo.es/download/geodisKey_14Jul04.pdf). NA = not applicable.

Clade	Inference chain	Inference
Clade 1-1	2-3-4-NO	Restricted gene flow with isolation by distance
Clade 1-2	2-3-5-6-7-8-YES	Restricted gene flow and isolation by distance with some long-distance dispersal over intermediate areas not occupied by the species
Clade 1-3	2-11-12-NO	Contiguous range expansion
Clade 1-4	2-3-4-NO	Restricted gene flow with isolation by distance
Clade 1-5	NA	No inference
Clade 1-6	2-3-5-6-7-YES	Restricted gene flow with isolation by distance with some long-distance dispersal
Clade 2-1	2-3-4-NO	Restricted gene flow with isolation by distance
Clade 2-2	2-3-4-NO	Restricted gene flow with isolation by distance
Total cladogram	2-11-17-4-NO	Restricted gene flow with isolation by distance

lation structure of *P. californica* consisted of both restricted gene flow with isolation by distance and long-distance dispersal events (Table 4). The total nested cladogram and clades 2-1, 2-2, 1-1, and 1-4 all provided inferences of restricted gene flow with isolation by distance (see Templeton et al. 1995 for a discussion of the statistical associations that produce each inference). The total cladogram and the 2-step clades represented the most ancient events that influenced population structure in this analysis (Posada and Crandall 2001). Clade 1-1 was an interior clade and represented more ancient historical processes. Clade 1-4 was a tip clade and was inferred to represent more recent historical processes (Templeton 1998).

The inference of restricted gene flow with isolation by distance at every level of the cladogram suggests that restricted gene flow has been a dominant mechanism influencing dispersal throughout the history of this organism. Clade 1-3 was an interior clade and was associated with a contiguous range expansion from the northeastern corner of the distribution to the west and south (Fig. 4). Clades 1-2 and 1-6 were tip clades. They both provided inferences of restricted gene flow with some long-distance dispersal. Clade 1-2 also identified a distinct southern lineage of the species that was defined by a single, nonsynonymous substitution.

Population genetic structure

The inferences from NCA suggested the existence of a northern *P. californica* lineage, char-

acterized by clades 1-3 and 1-6, and a genetically distinct southern lineage characterized by clade 1-2 (Figs 1, 4). Three of the 4 major haplotypes (haplotypes 10, 24, and 25) were found almost exclusively in the northern lineage. The 4th major haplotype (12), which had a single nonsynonymous substitution, defined the southern lineage and was also present in the northeastern part of the range of *P. californica*, where the 2 lineages overlap, but was absent from the rest of the northern lineage (Figs 1, 4). In addition, many minor haplotypes present in the northern lineage were absent from the southern lineage, and vice versa. The existence of 2 distinct lineages was further supported by the fact that the largest genetic difference between *P. californica* populations occurred between populations 1 (South Fork Boise River, Idaho) and 36 (Grays River, Wyoming). These populations are both in the Snake River Basin, but were genetically associated with the northern and southern lineages, respectively.

The northern lineage included an interior clade, but the southern lineage did not. This evidence from the NCA suggests that the northern lineage may be more ancient than the southern (Templeton 1998). Dispersal and diversification may have occurred in the north prior to the establishment of the southern lineage. These processes may have caused the higher mean genetic diversity observed in the northern lineage (0.00586 vs 0.00479). The region where the northern and southern lineages overlap had the highest mean genetic diversity (0.00640). High genetic diversity often is associated with the

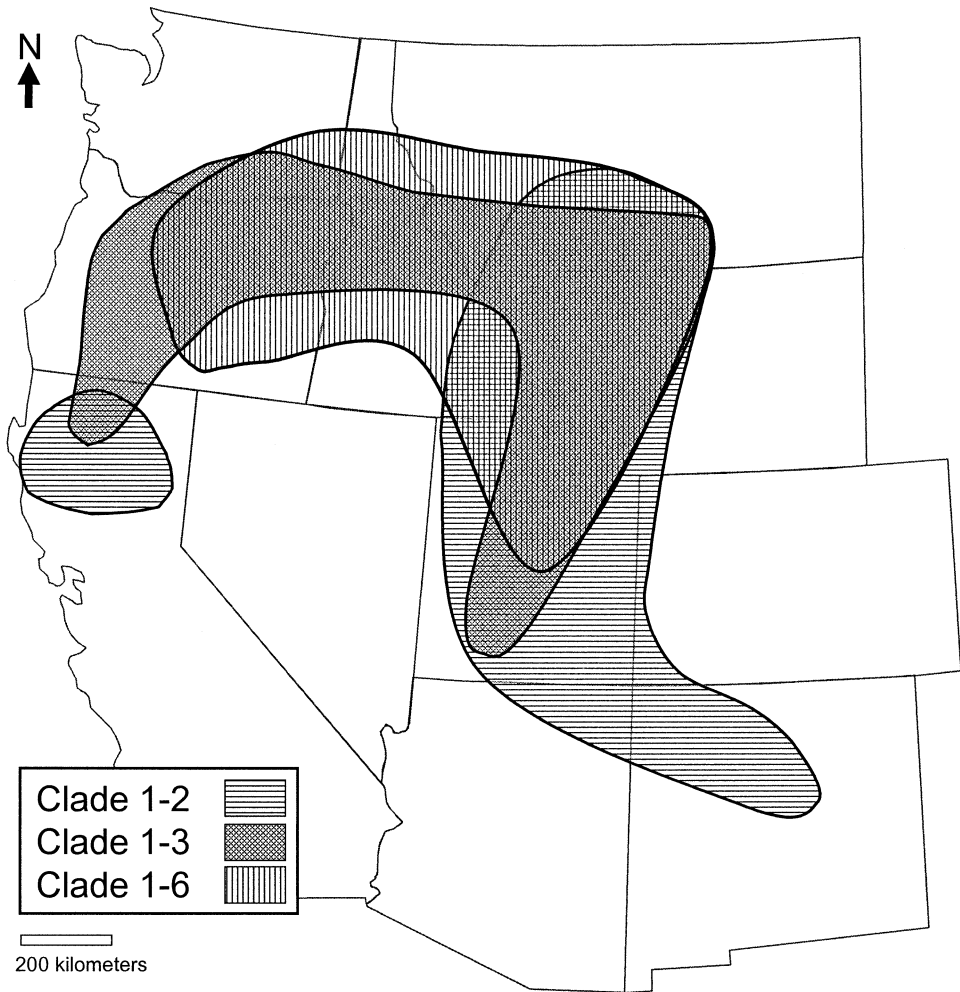


FIG. 4. Geographical distribution of clades 1-2, 1-3, and 1-6. In our study, clades 1-3 and 1-6 were designated as northern lineages, and clade 1-2 was designated as a southern lineage.

most ancestral part of a species' range, but the high diversity also could have been the result of secondary contact between 2 diverged lineages. Dispersal from a northeast (Wyoming–Montana) origin to the west and south is more concordant with the data than secondary contact because the phylogenetic tree suggested that the *P. californica*/*P. princeps* clade was derived from an eastern ancestor.

Means of dispersal

The inferences from our NCA suggest that the dominant mechanism in creating the pattern of genetic variation observed in *P. californica* was

restricted gene flow with isolation by distance. This mechanism is consistent with a pattern of downstream dispersal via river connections, and would have provided intermittent opportunities for dispersal as drainage connections changed through geological time. However, the geographic associations of some haplotypes in clades 1-2 and 1-6 cannot be explained by this mechanism. These associations suggest that direct gene flow between distant populations has occurred.

The inference of long-distance dispersal in clade 1-2 is based on the presence of a nonsynonymous substitution in population 21 (from California) and populations 2, 3, 7, 8, 9, 10, 13,

14, 28, 30, 32, 34, 36, and 40 (from New Mexico, Utah, Wyoming, Montana, and western Idaho) and the absence of the substitution in our samples from populations in the intermediate region (populations 1, 18, and 23 from Idaho, Utah, California, and Nevada, respectively). This substitution is the only nonsynonymous substitution in the data set. Its existence suggests that the inference of long-distance dispersal between these populations is not spurious. However, the statistical analysis performed by GeoDis cannot distinguish between long-distance dispersal and restricted gene flow with isolation by distance (if *P. californica* were once present in the intermediate region, but had since been extirpated).

No evidence exists to indicate that *P. californica* or other species that inhabit similar habitats (e.g., whitefish, cottids, and salmonids) were ever present in the southwestern Great Basin (Hubbs and Miller 1948). Instead, the aquatic organisms in this part of the Great Basin appear to have a southern origin (Hubbs and Miller 1948, Polhemus and Polhemus 2002). Nevertheless, whitefish, cottids, and salmonids are present in the Lahontan Basin, and rivers in the Lahontan Basin subunit of the Great Basin could have acted as dispersal routes, especially during pluvial times. *Pteronarcys californica* was collected in the Marys River, a major tributary to the upper Humboldt River, which flows west across the northeastern Lahontan Basin. Further investigation of other Lahontan Basin tributaries will be needed to determine whether the observed pattern of genetic variation is indicative of long-distance dispersal.

In clade 1–6, a statistically significant relationship was found between populations 27 and 20 (from the Payette River drainage and the Lost Rivers of Idaho) and populations 24 and 16 (from coastal Oregon). Five populations (1, 17, 19, 25, and 29) were sampled in the intermediate region (Boise River, John Day River, Salmon River, Columbia River; Fig. 1) but none of the populations exhibited any statistically significant relationships. This pattern of genetic variation is best explained by the occurrence of long-distance dispersal, i.e., overland dispersal of adult stoneflies.

Biogeography

The inferences of: 1) an origin in the north-eastern part of the range for *P. californica*, 2) both

westerly and southerly dispersal via river connections, and 3) occasional overland dispersal are consistent with the phylogenetic relationships of the 2 western species, which are derived from the more basal eastern taxa. If the ancestor of *P. californica* crossed the continental divide in the Miocene or Pliocene, then the likely source of the invasion would have been the Hudson Bay drainage via the upper Missouri River, which has been hypothesized to have been part of the Hudson Bay drainage at that time (Howard 1958, Smith 1987).

However, the relatively low level of mitochondrial differentiation seen in *P. californica* and the tendency of haplotypes to be broadly distributed indicate either high mobility or relatively recent (likely Pleistocene epoch) dispersal. During the Pleistocene, the Laurentian ice sheet established the modern path of the upper Missouri River. Low temperatures associated with glaciation could have facilitated the invasion of a *P. dorsata*-like ancestor from the upper midwest, across the plains, to the continental divide where the ancestral *P. californica*/*P. princeps* line was established. Subsequent dispersal probably allowed the divergence of the 2 species.

Given that *P. princeps* is present in isolated headwaters in the northern Great Basin, it seems plausible that this species moved down the upper Snake River and into the Great Basin, possibly through headwater transfers and short overland flights. A later, or parallel, westward expansion of the ancestral form could have established the northern lineage of *P. californica*. The relative timing of this invasion might have been resolved if specimens of *P. pictetii* had been included in the phylogenetic analysis. Unfortunately, we did not have access to that species. Regardless, this invasion could have followed the Snake River, but a more northerly path, perhaps along the upper Columbia River, Clark's Fork River, or upper Salmon River, also could have been the route taken by this species. Subsequently, another dispersal event from the eastern part of *P. californica*'s range occurred, this time to the south into the Bonneville, Colorado River, and Rio Grande drainages. This invasion would have established the southern lineage, all members of which share the same nonsynonymous substitution. *Pteronarcys californica* from the eastern and northern extremes of their range were not examined. Nymphs from both of these

areas should be examined to provide a complete picture of the overall patterns of dispersal.

In conclusion, the patterns of genetic variation that were observed in *P. californica* suggest that the species originated in the northeastern part of its present-day range and subsequently was divided into genetically distinct northern and southern lineages. Our data also suggested that downstream movement via river connections was probably the dominant mechanism of dispersal, but that long-distance dispersal also may have shaped the current pattern of genetic variation.

Acknowledgements

We thank K. Crandall, J. Marshall, R. Rader, B. Miller, and 2 anonymous referees for their critical reviews of this manuscript. We also thank A. Johnson, J. Crowley, M. McKell, D. Houston, G. Jacobi, and M. B. MacFarlane for assistance in collecting specimens of *P. californica*. B. Armitage, E. Dewalt, J. Earle, D. Lenat, S. Jordan, T. MacPherson, and B. J. Weibell kindly volunteered their time and assistance in collecting specimens of the other *Pteronarcys* species. B. Miller and D. Janetski helped perform the mtDNA analyses, and R. W. Baumann and C. R. Nelson verified identification of specimens.

Literature Cited

- BAUMANN, R. W., B. O. HUNTSMAN, AND B. C. KONDRATIEFF. 1999. Stoneflies (Plecoptera) of the Black Hills of South Dakota and Wyoming, USA: distribution and zoogeographic affinities. *Great Basin Naturalist* 59:1–17.
- BILTON, D. T., J. R. FREELAND, AND B. OKAMURA. 2001. Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics* 32:159–181.
- CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9:1657–1660.
- COLLIER, K. J., AND B. J. SMITH. 1998. Dispersal of adult caddisflies (Trichoptera) into forests alongside three New Zealand streams. *Hydrobiologia* 361: 53–65.
- CREASE, T. J., S. K. LEE, S. I. YU, K. SPITZE, N. LEHMAN, AND M. LYNCH. 1997. Allozyme and mtDNA variation in populations of the *Daphnia pulex* complex from both sides of the Rocky Mountains. *Heredity* 79:242–251.
- CRUZAN, M. B., AND A. R. TEMPLETON. 2000. Paleogeology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution* 15:491–496.
- FETZNER, J. W., AND K. A. CRANDALL. 2003. Linear habitats and the nested clade analysis: an empirical evaluation of geographic versus river distances using an Ozark crayfish (Decapoda: Cambaridae). *Evolution* 57:2101–2118.
- FREELAND, J. R., L. R. NOBLE, AND B. OKAMURA. 2000. Genetic diversity of North American populations of *Cristatella mucedo*, inferred from microsatellite and mitochondrial DNA. *Molecular Ecology* 9: 1375–1389.
- GRIFFITH, M. B., E. M. BARROWS, AND S. A. PERRY. 1998. Lateral dispersal of adult insects (Plecoptera, Trichoptera) following emergence from headwater streams in forested Appalachian catchments. *Annals of the Entomological Society of America* 91:195–201.
- HARTLEB, C. E., AND S. A. TIMM. 2000. Survival and hatching success of stonefly eggs (*Paragnetina media*) following ingestion by three stream fishes. *Journal of Freshwater Ecology* 15:107–114.
- HOUSEMAN, R. M., AND R. W. BAUMANN. 1997. Zoogeographic affinities of the stoneflies (Plecoptera) of the Raft River Mountains, Utah. *Great Basin Naturalist* 57:209–219.
- HOWARD, A. D. 1958. Drainage evolution in northeastern Montana and northwestern North Dakota. *Bulletin of the Geological Society of America* 69:575–588.
- HUBBS, C. L., AND R. R. MILLER. 1948. The Great Basin: with emphasis on glacial and post-glacial times. The zoological evidence. *Bulletin of the University of Utah* 38:18–166.
- HUGHES, J. M., P. B. MANTHER, A. L. SHELDON, AND F. W. ALLENDORF. 1999. Genetic structure of the stonefly, *Yoraperla brevis*, in populations: the extent of gene flow among adjacent montane streams. *Freshwater Biology* 41:63–72.
- JACKSON, J. K., AND V. H. RESH. 1989. Distribution and abundance of adult aquatic insects in the forest adjacent to a northern Californian stream. *Environmental Entomology* 18:278–283.
- KELLY, L. C., D. T. DILTON, AND S. D. RUNDLE. 2001. Genetic differentiation and dispersal in the Canary Island caddisfly *Mesophylax asperses* (Trichoptera: Limnephilidae). *Heredity* 86:370–377.
- KETMAIER, V., R. FOCHETTI, V. IANNILLI, AND E. DE-MATTHAEIS. 2001. Patterns of genetic differentiation and gene flow in central Italian populations of *Dinocras cephalotes* (Curtis, 1827) (Insecta, Plecoptera). *Archiv für Hydrobiologie* 150:457–472.
- KNOWLES, L. L., AND W. P. MADDISON. 2002. Statistical phylogeography. *Molecular Evolution* 11:2623–2635.
- KONDRATIEFF, B. C., AND R. W. BAUMANN (coordinators). 2000. Stoneflies of the United States. Northern Prairie Wildlife Research Center Home Page:

- <http://www.npwrc.usgs.gov/resource/distr/insects/sfly/sflyusa.htm> (version 26JUN2002). Northern Prairie Wildlife Research Center, Jamestown, North Dakota.
- KOVATS, Z. E., J. J. H. CIBOROWSKI, AND L. D. CORKUM. 1996. Inland dispersal of adult aquatic insects. *Freshwater Biology* 36:265–276.
- KUMAR, S., K. TAMURA, I. B. JAKOBSEN, AND M. NEI. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 12:1244–1245. (Available from: <http://www.megasoftware.net/>)
- MAGUIRE, B. J. 1963. The passive dispersal of small aquatic organisms and their colonization of isolated bodies of water. *Ecological Monographs* 33: 161–185.
- MERRITT, R. W., AND K. W. CUMMINS (EDITORS). 1996. An introduction to the aquatic insects of North America. 3rd edition. Kendall/Hunt, Dubuque, Iowa.
- MONAGHAN, M. T., P. SPAAK, C. T. ROBINSON, AND J. V. WARD. 2001. Genetic differentiation of *Baetis alpinus* Pictet (Ephemeroptera: Baetidae) in fragmented alpine streams. *Heredity* 86:395–403.
- MÜLLER, K. 1954. Investigations on the organic drift in north Swedish streams. Reports of the Institute for Freshwater Research, Drottingholm 35:133–148.
- NEI, M. 1972. Genetic distance between populations. *American Naturalist* 106:283–292.
- NELSON, C. H. 1988. Note on the phylogenetic systematics of the family Pteronarcyidae (Plecoptera), with a description of the eggs and nymphs of the Asian species. *Annals of the Entomological Society of America* 81:560–576.
- PENNINGER, M., AND D. POSADA. 2002. Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae: Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution* 56:1776–1788.
- POLHEMUS, D. A., AND J. T. POLHEMUS. 2002. Basins and ranges: the biogeography of aquatic true bugs (Insecta: Heteroptera) in the Great Basin. Pages 235–254 in R. Hershler, D. B. Madsen, and D. R. Currey (editors). *Smithsonian contributions to the earth sciences: Great Basin aquatic systems history*. Smithsonian Institution Press, Washington, DC.
- POSADA, D., AND K. A. CRANDALL. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* 16:37–45.
- POSADA, D., K. A. CRANDALL, AND A. R. TEMPLETON. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology* 9:487–488. (Inference key available from: http://darwin.uvigo.es/download/geodisKey_14Jul04.pdf)
- PYBUS, O. G., A. RAMBAUT, AND P. H. HARVEY. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1429–1437.
- RICKER, W. E. 1952. Systematic studies in Plecoptera. Indiana University Publications, Science Series 18: 1–200.
- RILEY, J. R., A. D. SMITH, D. R. REYNOLDS, A. S. EDWARDS, AND J. L. OSBOURNE. 1996. Tracking bees with harmonic radar. *Nature* 379:29–30.
- ROGERS, A. R., AND H. HARPENDING. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9:552–569.
- SCHULTHEIS, A. S., L. A. WEIGT, AND A. C. HENDRICKS. 2002. Gene flow, dispersal, and nested clade analysis, among populations, of the stonefly *Peltoperla tarteri* in the southern Appalachians. *Molecular Ecology* 11:317–327.
- SERVICE, M. E. 1993. Mosquito ecology field sampling methods. Elsevier Press, London, UK.
- SHELDON, A. L. 1979. Stonefly (Plecoptera) records from the basin and ranges of Nevada and Utah. *Great Basin Naturalist* 39:289–292.
- SMITH, G. R. 1987. Fish speciation in a western North American Pliocene rift lake. *Palaos* 2:436–445.
- STENSETH, N. C., AND W. C. LIDICKER. 1992. The study of dispersal: a conceptual guide. Pages 5–20 in N. C. Stenseth and W. C. Lidicker (editors). *Animal dispersal: small mammals as a model*. Chapman and Hall, New York/London, UK.
- STEWART, K. W., AND B. P. STARK. 1993. Nymphs of North American stonefly genera (Plecoptera). University of North Texas Press, Denton, Texas.
- TAYLOR, D. J., T. L. FINSTON, AND P. D. N. HEBERT. 1998. Biogeography of a widespread freshwater crustacean: pseudocongruence and cryptic endemism in the North American *Daphnia laevis* complex. *Evolution* 52:1648–1670.
- TAYLOR, D. W. 1985. Late Cenozoic history of the Pacific Northwest: evolution of freshwater drainages and molluscs in western North America. Pages 265–321 in C. J. Smiley (editor). *Late Cenozoic history of the Pacific Northwest*. Pacific Division American Association for the Advancement of Science and the California Academy of Science, San Francisco, California.
- TEMPLETON, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* 7:381–397.
- TEMPLETON, A. R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Evolution* 13:789–810.
- TEMPLETON, A. R., E. ROUTMAN, AND C. A. PHILLIPS. 1995. Separating population structure from population history: cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140:767–782.

- THOMAS, E. P., D. W. BLINN, AND P. KEIM. 1998. Do xeric landscapes increase genetic divergence in aquatic landscapes? *Freshwater Biology* 40:587–593.
- WATERS, T. F. 1972. The drift of stream insects. *Annual Review of Entomology* 17:253–272.
- WRIGHT, S. 1951. The genetical structure of populations. *Eugenics* 15:323–354.

Received: 26 June 2003

Accepted: 1 September 2004